Medical Secretaries' Registration Work in the Data-Driven Healthcare Era

Knudsen, Casper; Bertelsen, Pernille Scholdan

Published in:
Caring is Sharing – Exploiting the Value in Data for Health and Innovation

DOI (link to publication from Publisher):
10.3233/SHTI230115

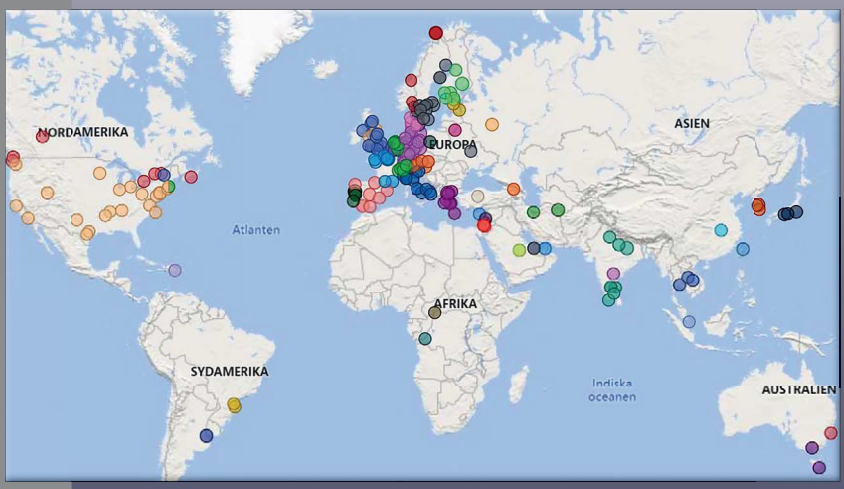
Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS 302

Caring is Sharing — Exploiting the Value in Data for Health and Innovation

Proceedings of MIE 2023



Editors:
Maria Hägglund
Madeleine Blusi
Stefano Bonacina
Lina Nilsson
Inge Cort Madsen
Sylvia Pelayo
Anne Moen
Arriel Benis
Lars Lindsköld
Parisis Gallos

IOS Press

Modern information and communication technologies make it easier for individuals to be involved in their own health and social care. They also facilitate contact between individuals and service providers and deliver more efficient tools for healthcare staff. Artificial Intelligence (AI) promises to bring even more benefits in the future, with more effectiveness and the provision of decision support.

This book presents the proceedings of the 33rd Medical Informatics Europe Conference, MIE2023, held in Gothenburg, Sweden, from 22 to 25 May 2023. The theme of MIE2023 was 'Caring is Sharing – Exploiting Value in Data for Health and Innovation', stressing the increasing importance of sharing digital-health data and the related challenges. The sharing of health data is developing rapidly, both in Europe and beyond, so the focus of the conference was on the enabling of trustworthy sharing of data to improve health. Topics covered include healthcare, community care, self-care, public health, and the innovation and development of future-proof digital-health solutions, and the almost 300 papers divided into 10 chapters also cover important advances in the sub domains of biomedical informatics: decision support systems, clinical information systems, clinical research informatics, knowledge management and representation, consumer health informatics, natural language processing, public health informatics, privacy, ethical and societal aspects among them.

Describing innovative approaches to the collection, organization, analysis, and data-sharing related to health and wellbeing, the book contributes to the expertise required to take medical informatics to the next level, and will be of interest to all those working in the field.

ISBN 978-1-64368-388-1 (print)
ISBN 978-1-64368-389-8 (online)
ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

CARING IS SHARING – EXPLOITING THE VALUE IN DATA FOR HEALTH AND INNOVATION

Studies in Health Technology and Informatics

Internationally, health informatics is driven by developments in biomedical technologies and medical informatics research that are advancing in parallel and form one integrated world of information and communication media and result in massive amounts of health data. These components include genomics and precision medicine, machine learning, translational informatics, intelligent systems for clinicians and patients, mobile health applications, data-driven telecommunication and rehabilitative technology, sensors, intelligent home technology, EHR and patient-controlled data, and Internet of Things.

The series Studies in Health Technology and Informatics (HTI) was started in 1990 in collaboration with EU programmes that preceded the Horizon 2020 to promote biomedical and health informatics research. It has developed into a highly visible global platform for the dissemination of original research in this field, containing more than 250 volumes of high-quality works from all over the world.

The international Editorial Board selects publications with relevance and quality for the field. All contributions to the volumes in the series are peer reviewed.

Volumes in the HTI series are submitted for indexing by MEDLINE/PubMed; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Google Scholar; Scopus; EMCare.

Volume 302

Recently published in this series

Vol. 301 B. Pfeifer, G. Schreier, M. Baumgartner and D. Hayn (Eds.), dHealth 2023 – Proceedings of the 17th Health Informatics Meets Digital Health Conference

Vol. 300 J. Mantas, A. Hasman and R. Haux (Eds.), Achievements, Milestones and Challenges in Biomedical and Health Informatics

Vol. 299 B. Blobel, B. Yang and M. Giacomini (Eds.), pHealth 2022 – Proceedings of the 19th International Conference on Wearable Micro and Nano Technologies for Personalized Health – 8–10 November 2022 Oslo, Norway

Vol. 298 P. Scott, J. Mantas, A. Benis, I. Ognjanović, K. Saranto, A. Ware, I. Wells and P. Gallos (Eds.), Digital Professionalism in Health and Care: Developing the Workforce, Building the Future – Proceedings of the EFMI Special Topic Conference 2022

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

Caring is Sharing – Exploiting the Value in Data for Health and Innovation

Proceedings of MIE 2023

Edited by

Maria Hägglund

Participatory eHealth and Health Data research group, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Madeleine Blusi

Department of Nursing, Umeå University, Umeå, Sweden

Stefano Bonacina

Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Lina Nilsson

eHealth Institute, Department of Medicine and Optometry, Linnaeus University, Kalmar, Sweden

Inge Cort Madsen

VIA University College, Aarhus, Denmark

Sylvia Pelayo

Clinical Investigation Center for Innovative Technologies (CIC-IT), Lille University, Lille, France

Anne Moen

Department of Public Health Science, University of Oslo, Oslo, Norway

Arriel Benis

Department of Digital Medical Technologies, Holon Institute of Technology, Holon, Israel

Lars Lindsköld

Region Västra Götaland, Gothenburg, Sweden

and

Parisis Gallos

Computational Biomedicine Research Lab, Department of Digital Systems, University of Piraeus, Greece

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-64368-388-1 (print)
ISBN 978-1-64368-389-8 (online)
Library of Congress Control Number: 2023936304
doi: 10.3233/SHTI302

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

Preface

The 33rd Medical Informatics Europe Conference, MIE2023, was held in Gothenburg, Sweden, from 22 to 25 May 2023. The Conference was hosted by the European Federation for Medical Informatics (EFMI) and organized by the Swedish Medical Informatics Association (SFMI). The Scientific Programme Committee was chaired by Associate Professor Maria Hägglund, Uppsala University and Uppsala University Hospital.

The overarching theme of MIE2023 was "Caring is Sharing – Exploiting Value in Data for Health and Innovation", stressing the increasing importance of sharing digital-health data and the challenges related to this. The theme is closely connected to the rapid development of health-data sharing in Europe and globally, so the focus was on the opportunities provided by health informatics and research to enable the trustworthy sharing of health data to improve human health. This includes healthcare, community care, self-care, public health, and the innovation and development of future-proof digital health solutions. Modern information and communication technologies make it easier for individuals to be involved in their own health and social care, facilitate contact between individuals and service providers, and provide more efficient tools for healthcare staff. Furthermore, artificial intelligence (AI) promises to be of benefit in the future, bringing more effectiveness in some situations and providing decision support.

The COVID-19 pandemic has not only increased the speed of implementation and adoption of eHealth throughout Europe and globally, but has also highlighted how weak infrastructure and obstructive data-sharing regulations can hinder effective public health interventions and innovation. The European Union has highlighted this in the European Health Data Space (EHDS) legislation proposal. The EHDS will require all EU member states to step up their digitalization of healthcare, for both the primary and secondary use of health data. These proceedings contribute to this work, providing the expertise and cutting-edge research required to implement the EHDS proposal.

Throughout this publication, readers will find innovative approaches to the collection, organization, analysis, and especially the sharing of data and knowledge related to health and wellbeing. Included papers also cover important advances in the sub domains of biomedical informatics; decision support systems, clinical information systems, clinical research informatics, knowledge management, and representation, consumer health informatics, natural language processing, public health informatics, privacy, ethical and societal aspects, etc.

The Proceedings are published as an e-book, with open access for ease of use and browsing without any loss of the advantages of indexing and citation, in the biggest Scientific Literature Databases, such as Medline and Scopus provided by the series of Studies in Health Technology and Informatics (HTI) of IOS Press.

The Editors,
Maria Hägglund, Madeleine Blusi, Stefano Bonacina, Lina Nilsson, Inge Cort Madsen, Anne Moen, Lars Lindsköld, Arriel Benis, Parisis Gallos

Uppsala, 02.04.2023

This page intentionally left blank

About the Conference

The Conference

From May 22nd to 25th 2023, the European Federation for Medical Informatics (EFMI) organized the 33rd Medical Informatics Europe Conference (MIE2023) in Gothenburg, Sweden. The conference was managed by the Swedish Association of Medical Informatics (SFMI, https://www.sfmi.se/) together with the Vitalis conference (https://vitalis.nu/) and Svenska Mässan (The Swedish Exhibition and conference center, https://en.svenskamassan.se/).

EFMI (https://efmi.org/) is the leading organization in medical informatics in Europe, founded in 1976, and represents 30 countries through their respective national health informatics associations. EFMI is organized as a not-for-profit organization concerned with the theory and practice of Information Science and Technology within Healthcare and Health Sciences in a European context.

MIE is a medical informatics conference that aims at promoting research and development in biomedical and health informatics. The conference proposes scientific sessions consisting of oral presentations of peer-reviewed full papers and short communication papers. The conference also includes panels, workshops, demos, and tutorials, some of them being prepared by EFMI working groups. A large exhibition of peer-reviewed posters is also part of the conference.

In 2023, the special theme of the conference was “Caring is Sharing - Exploiting Value in Data for Health and Innovation” (https://www.mie2023.org/).

Conference Topics include (but are not limited to):

- Special Topic: Caring is Sharing – exploiting value in data for health and innovation
- Bioinformatics
- Citizen health informatics
- Decision support
- Education
- Health information systems
- Human Factors and organisational issues
- Knowledge and Information representation and modelling
- Medical Robotics
- Natural Language Processing
- Patient records
- Public Health and Epidemiology Informatics
- Security and Safety
- Sensors, signals, and Imaging Informatics
- Societal aspects
- Telehealth
- Visualization

The MIE2023 conference included four keynotes by internationally recognized experts in medical informatics:

- **Tom Lawry**, a strategic advisor to health leaders worldwide and the best-selling author of Hacking Healthcare – How AI and the Intelligence Revolution will Reboot an Ailing System. He is the Managing Director of Second Century Technology and a former Microsoft executive where he served as National Director of AI for Health and Life Sciences, Director of Worldwide Health, and Director of Organizational Performance for the company's first health incubator. Prior to Microsoft, Tom was a Senior Director at GE Healthcare, the founder of two venture-backed healthcare software companies, and a health system executive. Tom Lawry will give a joint keynote for Vitalis and MIE2023.
- **Professor Dipak Kalra,** President of the European Institute for Innovation through Health Data (i~HD) and an internationally renowned expert in health informatics. With a distinguished career in healthcare and technology, Professor Kalra has contributed significantly to the advancement of electronic health records (EHRs), interoperability standards, and patient data privacy. His extensive expertise encompasses clinical data modeling, patient-centered health records, and health data governance. In his keynote address, Professor Dipak Kalra shares his insights on the critical role of health data in driving innovation and improving patient care. His talk covers essential topics such as the ethical use of health data, data sharing across borders, and the challenges and opportunities in implementing new health informatics standards.
- **Dana Lewis** founded the open-source artificial pancreas movement (known as "OpenAPS"), working to make safe and effective automated insulin delivery (AID) technology available (sooner) for people with diabetes around the world for the past 8 years. She authored the book, *'Automated Insulin Delivery: How artificial pancreas "closed loop" systems can aid you in living with diabetes'*, to help more people understand automated insulin delivery systems. Her peer-reviewed publications have been cited more than 1,300 times. She has collected numerous types of individual data and conducted research with it, while also working to support communities of patients and medical and academic researchers to harness the power of real-world shared data for improving healthcare.
- **Persephone Doupi**, Senior Medical Officer at Finnish Institute for Health and Welfare (THL). Persephone shares her insights about the European Union Data Strategy that aims to create shared European data infrastructures that allow for cross-border exchange and availability of high-quality data. The European Health Data Space is the first of these infrastructures and focuses on using health and social data to advance public health, research, and innovation.

Scientific Program Committee / Chair and co-chairs (alphabetic)

- Associate Prof. Maria Hägglund (chair), Uppsala University and Uppsala University Hospital, Uppsala MedTech Science and Innovation, Department of Women's and Children's Health, Uppsala, Sweden
- Associate Prof. Madeleine Blusi (co-chair), Umeå University, Department of Nursing, Umeå, Sweden

- Dr. Stefano Bonacina (co-chair), Karolinska Institutet, Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Stockholm, Sweden
- Associate Prof. Lina Nilsson (co-chair), Linnaeus University, eHealth Institute, Department of Medicine and Optometry, Kalmar, Sweden
- RN, MI. Inge Cort Madsen (co-chair), VIA University College, Aarhus, Denmark
- Prof. Sylvia Pelayo (co-chair), Lille University, Clinical Investigation Center for Innovative Technologies (CIC-IT) 1403, Lille, France
- Prof. Anne Moen (co-chair), University of Oslo, Department of Public Health Science, Oslo, Norway
- Dr. Lars Lindsköld (co-chair), Region Västra Götaland, Gothenburg, Sweden

Editorial Committee / Co-chairs (alphabetic)

- Associate Prof. Arriel Benis (co-chair), Department of Digital Medical Technologies, Holon Institute of Technology, Holon, Israel
- Dr. Parisis Gallos (co-chair), Computational Biomedicine Research Lab, Department of Digital Systems, University of Piraeus, Greece.

Peer Review Process

We received over 420 submissions from 49 countries. A thorough reviewing process was conducted with valuable support from 260 active reviewers. About all submissions were reviewed by at least three reviewers and assessed by one SPC co-chair. Based on their recommendations, final decisions were made by SPC members during a two-day face-to-face meeting in Gothenburg, Sweden. Papers requiring major revisions underwent another review by SPC members.

Finally, among the 265 full papers, 51 short communication papers, 48 posters, 12 demonstrations, 13 panels, 28 workshops, and 4 tutorials submitted, 156 papers (acceptance rate of 59%), 40 short communication papers (conversion of 16 full papers), 97 posters (conversion of 46 full papers, and 13 short communications), 10 demonstrations, 13 panels, 25 workshops, and 4 tutorials were accepted. All accepted full papers, short communication papers, and posters are included in these proceedings.

We want to thank the SPC co-chairs and all the reviewers for their invaluable contributions to MIE2023.

Main Sponsors (as of April 16th, 2023)

Region Västra Götaland
Cambio Healthcare Systems

Addresses for correspondence:
Maria.Hagglund@uu.se and Lars.Lindskold@SFMI.se

Reviewers

Mr. Karthik Adapa
Ms. Rose-Mharie Åhlfeldt
Ms. Karin Ahlin
Ms. Leila Ahmadian
Mr. Xavie Aimé
Mr. Tanvir Alam
Ms. Marta Alić
Mr. Samer Alkarkoukly
Mr. Yasser Alsafadi
Ms. Sahar Alsharif
Ms. Shaymaa Al-Shukri
Ms. Elske Ammenwerth
Ms. Elisavet Andrikopoulou
Mr. Alırıza Arıbaş
Mr. Kfeel Arshad
Mr. František Babič
Mr. Mansoor Baig
Ms. Gabriella Balestra
Mr. Charalampos Balis
Mr. Alireza Banaye Yazdipour
Ms. Franziska Bathelt
Mr. Patrice Bellot
Mr. Mohamed Ben Said
Mr. Arriel Benis
Mr. Patrick Bergman
Mr. Joao H Bettencourt-Silva
Mr. Geraldo Bezerra da Silva Junior
Ms. Jen Bichel-Findlay
Mr. Jonas Bienzeisler
Mr. Antonis Billis
Ms. Tetiana Biloborodova
Mr. Bernd Blobel
Ms. Madeleine Blusi
Mr. Stefano Bonacina
Mr. Pietro Bosoni
Mr. Oliver J. Bott
Mr. Tobias Brix
Ms. Laura Broasca
Mr. Thomas Brox Røst
Mr. Jean Charlet
Ms. Helen Chen
Mr. ching lung cheng
Ms. Andja Cirkovic
Ms. Beatrice Coldewey
Mr. Ronald Cornet
Mr. Ricardo Correia
Mr. Sébastien Cossin
Mr. Carlos Costa
Ms. Melanie Courtine
Ms. Mihaela Crișan - Vida
Mr. Jaime Delgado
Ms. Kerstin Denecke
Mr. Thomas Deserno
Ms. Verena Dornauer
Ms. Cora Marisa Drenkhahn
Mr. Martin Dugas
Mr. Urs Eisenmann
Mr. Eric Eisenstein
Mr. Peter Ellison
Mr. Florian Endel
Mr. Jean-Baptiste Excoffier
Mr. Thibaut Fabacher
Mr. Göran Falkman
Ms. Mina Fallah
Ms. Razieh Farrahi
Mr. Mohammad Faysel
Mr. Georgios Feretzakis
Mr. Joseph Finkelstein
Mr. Mircea Focsa
Ms. Marina Fotteler
Ms. Maite Franco
Mr. Alberto Freitas
Mr. Gerard Freriks
Ms. Eliana Ludmila Frutos
Ms. Kate Fultz Hollis
Mr. Jan Gaebel
Mr. Parisis Gallos
Mr. Julián García-García
Ms. Gersende Georg
Mr. Spiros Georgakopoulos
Mr. Charalabos Georgiadis
Ms. Raheleh Ghouchan Nezhad Noor Nia
Mr. Mauro Giacomini
Ms. Noemi Giordano
Ms. Carola Gißke
Mr. Robin Glauser
Ms. Natalia Grabar
Mr. Vincenzo Guardabasso
Mr. Sergio Guinez-Molinos
Mr. Charles Gutteridge
Mr. Werner O Hackl
Ms. Maria Hägglund
Ms. Josefin Hagström
Ms. Sadrieh Hajesmaeel Gohari
Ms. Tora Hammar
Mr. Eisuke Hanada
Ms. Janna Hastings
Mr. Yaseen Hayajneh
Mr. Dieter Hayn
Mr. Harald Heinzl
Ms. Mira Hercigonja-Szekeres
Ms. Asmâa hidki
Mr. Jacob Hofdijk
Mr. Felix Holl
Ms. Ursula Hübner
Mr. Jens Hüsers
Mr. Miron Iancu
Mr. Mohaimenul Islam
Ms. Franziska Jahn
Mr. Markus Jochem
Ms. Pauline Johansson
Mr. Owen Johnson
Mr. Charles Kahn
Ms. Christine Kakalou
Mr. Athanasios Kallipolitis
Ms. Bridget Kane
Mr. Haralampos Karanikas
Mr. Konstantinos Katzis
Mr. Oladimeji Kazeem
Ms. Anna Kharko
Mr. Santisith Khiewkhern
Mr. G Kim
Ms. Ulla-Mari Kinnunen
Mr. Athanasios Kiourtis
Ms. Petra Knaup
Ms. Ann-Kristin Kock-Schoppenhauer
Mr. Miroslav Končar
Mr. Appoh Kouamé
Mr. Dionysios Koulouris
Mr. Frank Kramer
Mr. Markus Kreuzthaler
Mr. Jörn Krückeberg

Mr. Pushpa Kumarapeli
Mr. Tomohiro Kuroda
Ms. Ekaterina Kutafina
Mr. Hallvard Lærum
Mr. Stefan Lagrosen
Mr. Antoine Lamer
Mr. Jean-Baptiste Lamy
Mr. Alberto Lavelli
Ms. Eunji Lee
Ms. Agnieszka Lemanska
Ms. Pashalina Lialiou
Mr. Joseph Liaskos
Ms. Valentina Lichtner
Mr. Lars Lindsköld
Mr. Matthias Löbe
Ms. Julia Logan
Mr. Guillermo Lopez-Campos
Mr. Chi-Hua Lu
Ms. Diana Lungeanu
Ms. Andriana Magdalinou
Mr. Ilias Maglogiannis
Ms. Marianne Maktabi
Mr. Sadouanouan MALO
Mr. Patrick Mangesius
Mr. John Mantas
Mr. Romaric Marcilly
Ms. Aniek Markus
Mr. Michael Marschollek
Mr. Jakir Hossain Bhuiyan Masud
Ms. Argyro Mavrogiorgou
Mr. Enkeleint A. Mechili
Mr. George Mihalas
Ms. Iskra Mircheva
Ms. Atiehsadat Mirkhani
Mr. Rakshit Mittal
Ms. Anne Moen
Ms. Reyhaneh Mohammadi
Ms. Maryati Mohd Yusof
Mr. Carlos Molina
Mr. Jonas Moll
Mr. Dominik Müller
Mr. Petteri Mussalo
Mr. Eustache Muteba Ayumba
Mr. Amin Naemi
Mr. Pantelis Natsiavas
Ms. Laura Naumann
Ms. Sarah Nee
Ms. Lina Nilsson
Ms. Evalill Nilsson
Ms. Ivana Ognjanovic
Mr. Antonios Pardos
Mr. Carlos Luis Parra-Calderón
Mr. Fabrizio Pecoraro
Mr. Niels Peek
Ms. Sylvia PELAYO
Ms. Laura-Maria Peltonen
Mr. Michele Persiani
Ms. Carolyn Petersen
Ms. Linda Peute
Ms. Jessica Pinaire
Mr. Lembit Pirn
Ms. Gouri Prakash
Ms. Mirela Prgomet
Mr. Michael Prinz
Mr. Vytenis PUNYS
Mr. Manuel Quesada-Martínez
Ms. Maria Qvarfordt
Mr. Jean Louis Raisaro
Mr. Nadav Rappoport
Mr. Giuseppe Rauch
Mr. Daniel Ray
Mr. Niklas Reimer
Ms. Tanja Rejc
Ms. Sara Riggare
Ms. Ana Rocha
Ms. Nancy Roderer
Mr. Alejandro Rodríguez González
Mr. Rainer Röhrig
Ms. Anna Rossander
Ms. Oksana Ryabinina
Mr. Etienne Saliez
Mr. Michail Sarafidis
Ms. Kaija Saranto
Mr. Stefan Sauermann
Mr. André Scherag
Mr. Jan Schladetzky
Mr. Stephan Schug
Mr. Marco Schweitzer
Ms. Karima Sedki
Ms. Danielle Sent
Ms. Brigitte Sseroussi
Mr. Syed-Abdul Shabbir
Ms. Mary Sharp
Ms. Toria Shaw Morawski
Mr. Kirubel Biruk Shiferaw
Mr. Michael Shifrin
Ms. Marieke Sijm-Eeken
Ms. Inna Skarga-Bandurova
Ms. Georgia Sovatzidi
Mr. Nicolai Spicher
Mr. Dalibor Stanimirovic
Mr. Milton Stern
Mr. Vasile Stoicu-Tivadar
Ms. Lacramioara Stoicu-Tivadar
Ms. Jenny Stritzel
Mr. Antonis Stylianides
Ms. Hanna Suominen
Mr. Stephen Swift
Mr. Oscar Tamburis
Ms. Sylvia Thun
Mr. Lucas Triefenbach
Ms. Maria Tsirintani
Ms. Niruwan Turnbull
Ms. Melina Tziomaka
Mr. Adrien UGON
Mr. Philipp Urbauer
Mr. William Van Woensel
Ms. Sidsel Villumsen
Mr. Thomas Vollmer
Ms. Julia S. Volmerg
Ms. Hanna von Gerich
Mr. Maxime Wack
Mr. Zhan Wang
Mr. Patrick Weber
Mr. Joshua Wiedekopf
Ms. Marisa Wilson
Mr. Alfred Winter
Mr. Dominik Wolff
Mr. Dimitrios Zarakovitis
Mr. Emmanouil Zoulias

This page intentionally left blank

Contents

Preface v
Maria Hägglund, Madeleine Blusi, Stefano Bonacina, Lina Nilsson, Inge Cort Madsen, Anne Moen, Lars Lindsköld, Arriel Benis, Parisis Gallos

About the Conference vii

Section 1. Caring is Sharing – Exploiting Value in Data for Health and Innovation

German Claims Data for Real-World Research: Content Coverage Evaluation in OMOP CDM 3
Elisa Henke, Michéle Zoch, Ines Reinecke, Melissa Spoden, Thomas Ruhnke, Christian Günster, Martin Sedlmayr and Franziska Bathelt

Can Patient Contributed Data (PCD) Leverage Connected Health Technology for Cardiac Rehabilitation in Austria? 8
Rada Hussein, Adrienne Pichon, Jan Oldenburg, Mahdi Sareban and Josef Niebauer

Towards a Clinically Meaningful Model to Structure the Development of Interoperable Order Sets, Applicable to the Point of Care in Any EMR 13
Mobin Yasini, Dennis Rausch, Lise Marin, Ingrid Hochheim, Navinder Singh Dhillon and Michael Dahlweid

Application of Process Mining for Modelling Small Cell Lung Cancer Prognosis 18
Luca Marzano, Sebastiaan Meijer, Asaf Dan, Salomon Tendler, Luigi De Petris, Rolf Lewensohn, Jayanth Raghothama and Adam S. Darwich

Participatory Study to Explore Healthcare Professionals' Perceptions of a Connected Digital Solution for Adherence Monitoring of Recombinant Human Growth Hormone Treatment: Study Protocol and First Findings 23
Octavio Rivera-Romero, Maria Felicia Faienza, Edoardo Vergani, Chantal Zucchiatti and Ekaterina Koledova

Utility-Preserving Anonymization in a Real-World Scenario: Evidence from the German Chronic Kidney Disease (GCKD) Study 28
Lisa Pilgram, Elke Schäffner, Kai-Uwe Eckardt, Fabian Prasser and GCKD Investigators

The Necessity of Multiple Data Sources for ECG-Based Machine Learning Models 33
Lucas Plagwitz, Tobias Vogelsang, Florian Doldi, Lucas Bickmann, Michael Fujarski, Lars Eckardt and Julian Varghese

Predicting Progression of Type 2 Diabetes Using Primary Care Data with the Help of Machine Learning 38
Berk Ozturk, Tom Lawton, Stephen Smith and Ibrahim Habli

Breaking Barriers for Interoperability: A Reference Implementation of CSV-FHIR Transformation Using Open-Source Tools 43
Samer Alkarkoukly, Md. Mostafa Kamal and Oya Beyan

Implementation Status of the Proposal for a Regulation of the European Health Data Space in Portugal: Are We Ready for It? 48
Margarida Mateus, Maria Loureiro, A. Raquel Fernandes, Miguel Oliveira and Ricardo Cruz-Correia

Health Synthetic Data to Enable Health Learning System and Innovation: A Scoping Review 53
Shu-Feng Tsao, Kam Sharma, Hateem Noor, Alan Forster and Helen Chen

Reproducibility in 2023 – An End-to-End Template for Analysis and Manuscript Writing 58
Jonathan M. Mang, Hans-Ulrich Prokosch and Lorenz A. Kapsner

Assessing the FAIRness of Deep Learning Models in Cardiovascular Disease Using Computed Tomography Images: Data and Code Perspective 63
Kirubel Biruk Shiferaw, Atinkut Zeleke and Dagmar Waltemath

No Transfer Without Validation: A Data Sharing Framework Use Case 68
Hauke Hund, Reto Wettstein, Christopher Hampf, Martin Bialke, Maximilian Kurscheidt, Simon T. Schweizer, Christoph Zilske, Simon Mödinger and Christian Fegeler

Comparative Analysis of Electrodermal Activity Decomposition Methods in Emotion Detection Using Machine Learning 73
Sriram Kumar P, Praveen Kumar Govarthan, Nagarajan Ganapathy and Jac Fredo Agastinose Ronickom

Mapping Korean National Health Insurance Claim Codes for Laboratory Test to SNOMED CT 78
Sumi Sung, Hyeoun-Ae Park, Seul Ki Park, Hyesil Jung, Hannah Kang and Min Sun Lee

WHO's Community-Centered Epidemic and Pandemic Information Platform: Hive 83
Sylvie Briand, Sarah Hess, John Lee, Katherine Sheridan, Tim Nguyen and Brian Yau

Automatic Outlier Detection in Laboratory Result Distributions Within a Real World Data Network 88
Aida Muñoz Monjas, David Rubio Ruiz, David Pérez-Rey and Matvey Palchuk

The Importance of Being FAIR and FAST – The Clinical Epidemiology and Study Platform of the German Network University Medicine (NUKLEUS) 93
Dagmar Krefting, Gabi Anton, Irina Chaplinskaya-Sobol, Sabine Hanss, Wolfgang Hoffmann, Sina M. Hopff, Monika Kraus, Roberto Lorbeer, Bettina Lorenz-Depiereux, Thomas Illig, Christian Schäfer, Jens Schaller, Dana Stahl, Heike Valentin, Peter Heuschmann and Janne Vehreschild

Data Sharing Platform for MIMIC-IV and MIMIC-ED Data Marts: Designing a Data Retrieving System Based on the Intra-Hospital Patient Transfer Pathway 98
Sharareh Rostam Niakan Kalhori, Thomas M. Deserno, Jamal Soleiman and Shayan Kasiri Habibabadi

The Need for a Non-Invasive Technology for Endometriosis Detection and Care 103
Ariane Hine, Juliana Bowles and Thais Webber

An Annotation Workbench for Semantic Annotation of Data Collection Instruments 108
Julia Sasse and Juliane Fluck

Implementation of HL7 FHIR-Based Interoperability Profiles to Manage Care Plans for Multimorbid Patients with Mild Dementia 113
Mert Gencturk, A. Emre Akpinar and Gokce B. Laleci Erturkmen

Towards Prediction of Injuries in Traffic Accidents 118
Viktor M.G. Sobotta, Maximilian Flormann, Roman Henze and Thomas M. Deserno

Predicting Family Implementation of Complementary and Alternative Medicine in Autism Online Communities 123
Zhenni Ni, Yuxing Qian, Hao Li and Zhizhen Yao

Integrating Cross-Departmental Data 125
Julia Gehrmann and Oya Beyan

Classification of Parkinson's Disease from Voice – Analysis of Data Selection Bias 127
Alexander Brenner, Catharina Marie Van Alen, Lucas Plagwitz and Julian Varghese

Does Using a Stacking Ensemble Method to Combine Multiple Base Learners Within a Database Improve Model Transportability? 129
Cynthia Yang, Egill A. Fridgeirsson, Jan A. Kors, Jenna M. Reps, Peter R. Rijnbeek, Jenna Wong and Ross D. Williams

Implementation of the Regulation on the European Health Data Space (EHDS) in Croatia 131
Emanuel Brađašević

Mapping the SPHN Dataset to FHIR 133
Sophie Anne Inès Klopfenstein, Sylvia Thun, Katrin Crameri and Caroline Stellmach

Assessing Quality of Life Using FHIR – How to Combine Patient Reported Outcome with Patient Generated Data for Better Compliance 135
Chantal N.L. Beutter, Patrick Werner, Stefan Sigle, Uwe Marc Martens and Christian Fegeler

Portal of Medical Data Models: Application in Federated Data Capture 137
Matthias Ganzinger, Max Blumenstock, Christian Niklas and Martin Dugas

The DELPHI Library: Improving Model Validation, Transparency and Dissemination Through a Centralised Library of Prediction Models 139
Ross D. Williams, Sicco den Otter, Jenna M. Reps and Peter R. Rijnbeek

Continuity of Patient Information to Palliative Care 141
Anne Kuusisto, Kaija Saranto, Päivi Korhonen and Elina Haavisto

Secure Sharing of Health-Related Data: Research Description of the VINTER, DELFIN, and HEIDA Projects 143
Helena M. Linge and Rickard Brännvall

Improving Healthcare Quality with an LHS 145
Ricardo L. Santos and Ricardo Cruz-Correia

The MeDaX Knowledge Graph Prototype 147
Judith A.H. Wodke, Lea Michaelis and Ron Henkel

A Toolchain for Big Data Analyses in the Intelligent Cognitive Operating Room 149
Elena Kirsch, Denise Junger, Christian Kücherer and Oliver Burgert

Classification of Healthcare Professionals 151
Annina Maag, Renate Ranegger and Dieter Baumberger

An Enhanced Standardization and Qualification Mechanism for Heterogeneous Healthcare Data 153
George Manias, Ainhoa Azqueta-Alzúaz, Andrea Damiani, Eshita Dhar, Eleftheria Kouremenou, Marta Patino-Martínez, Mariachiara Savino, Syed Abdul Shabbir and Dimosthenis Kyriazis

Section 2. Health Information Systems, Patient Records, Visualization Tools and Security & Safety

Towards Safe Conversational Agents in Healthcare 157
Kerstin Denecke

Development of a Command Line Interface for the Analysis of Result Sets from Automated Queries to Literature Databases 162
Tobias J. Brix, Malte Berentzen, Ludger Becker, Michael Storck and Julian Varghese

Cross-Registry Benchmarking of Data Quality: Lessons Learned 167
Jürgen Stausberg, Sonja Harkener, Christoph Engel, Robert Finger, Carsten Heinz, Ekkehart Jenetzky, Patrick Jersch, David Martin, Rüdiger Rupp, Martin Schoenthaler, Barbara Suwelack and Jeannine Wegner

'Quer N0 AVC' for Monitoring Stroke Patients' Healthcare Using a Mobile App 172
Gabrielle Dos Santos Leandro, Ivonei Bittencourt, Vivian Nagel, Carla Heloisa Cabral Moro and Claudia Moro

Enabling Clinical Trials of Artificial Intelligence: Infrastructure for Heart Failure Predictions 177
Arian Ranjbar, Jesper Ravn, Erland Ronningen and Ole Hanseth

Post Hoc Sample Size Estimation for Deep Learning Architectures for ECG-Classification 182
Lucas Bickmann, Lucas Plagwitz and Julian Varghese

How to Represent the Patient Voice in the Electronic Health Record? 187
Line Silsand, Gro Hilde Severinsen, Gro R. Berntsen and Aslak Steinsbekk

User Satisfaction with Recently Deployed Electronic Health Records 192
Ove Lintvedt, Luis Marco-Ruiz and Rune Pedersen

Conception and Development of a Targeted Alert System: Multisystem Considerations 197
Frederic Ehrler and Katherine Blondon

Why Are Data Missing in Clinical Data Warehouses? A Simulation Study of How Data Are Processed (and Can Be Lost) 202
Sonia Priou, Guillaume Lame, Marija Jankovic, Gilles Chatellier, Romain Bey, Christophe Tournigand, Christel Daniel and Emmanuelle Kempf

Nationally Shared Medication Lists – Describing Systems in the Nordic Countries 207
Tora Hammar, Trine S. Bergmo, Cille Bülow, Stine S. Clausen, Unn Sollid Manskow, Johanna Timonen and Anette Vik Jøsendal

An Architecture for Providing Personalized Digital Health 212
Jaime Delgado and Silvia Llorente

Social Determinants of Health Data Quality at Different Levels of Geographic Detail 217
Melody L. Greer, Maryam Y. Garza, Steve Sample and Sudeepa Bhattacharyya

Development of OSOMO Prompt Mobile Application on Elderly Population for Village Health Volunteers Using the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) Model 222
Niruwan Turnbull, Kukiat Tudpor, Siwaporn Chimphoklang, Sureewan Siladlao, Nutchanun Kamonroek and Kannikar Hannah Tatiyaworawattanakul

Implementation, Adoption and Use of the Kanta Services in Finland 2010–2022 227
Vesa Jormanainen, Tuulikki Vehko, Marina Lindgren, Ilmo Keskimäki and Minna Kaila

Optimization of Pre-Ictal Interval Time Period for Epileptic Seizure Prediction Using Temporal and Frequency Features 232
Abdul Aleem Shaik Gadda, Dhanvi Vedantham, John Thomas, Yuvaraj Rajamanickam, Ramshekhar N. Menon and Jac Fredo Agastinose Ronickom

DeepTSE: A Time-Sensitive Deep Embedding of ICU Data for Patient Modeling and Missing Data Imputation 237
Michael Fujarski, Christian Porschen, Lucas Plagwitz, Daniel Stroth, Catharina Marie Van Alen, Mahan Sadjadi, Raphael Weiss, Alexander Zarbock, Thilo Von Groote and Julian Varghese

Adolescents Identifying Errors and Omissions in Their Electronic Health Records: A National Survey 242
Josefin Hagström, Charlotte Blease, Anna Kharko, Isabella Scandurra and Maria Hägglund

The Smart Data Extractor, a Clinician Friendly Solution to Accelerate and Improve the Data Collection During Clinical Trials 247
Sophie Quennelle, Maxime Douillet, Lisa Friedlander, Olivia Boyer, Antoine Neuraz, Anita Burgun and Nicolas Garcelon

Monitoring Distributed Business Processes in Biomedical Research 252
Christoph Zilske, Maximilian Kurscheidt, Simon T. Schweizer, Hauke Hund, Simon Mödinger and Christian Fegeler

Low Valence Low Arousal Stimuli: An Effective Candidate for EEG-Based Biometrics Authentication System 257
Jahanvi Jeswani, Praveen Kumar Govarthan, Abirami Selvaraj, Amalin Prince, John Thomas, Mohanavelu Kalathe, Sreeraj V S, Venkat Subramaniam and Jac Fredo Agastinose Ronickom

Medical Secretaries' Registration Work in the Data-Driven Healthcare Era 262
Casper Knudsen and Pernille Bertelsen

HEIDA: Software Examples for Rapid Introduction of Homomorphic Encryption for Privacy Preservation of Health Data 267
Rickard Bränvall, Henrik Forsgren and Helena Linge

Identifying Relevant FHIR Elements for Data Quality Assessment in the German Core Data Set 272
Christian Draeger, Erik Tute, Carsten Oliver Schmidt, Dagmar Waltemath, Martin Boeker, Alfred Winter and Matthias Löbe

Automating Structured Results Communication to Expedite Imaging-Directed Care in Spine Oncology 277
Alvand Hassankhani, Raghav R. Mattay, Colbey W. Freeman and Charles E. Kahn Jr.

Greek Hospital Data Mining and Analysis 282
Maria Chintiroglou, Haralampos Karanikas and Sotiris Tasoulis

Challenges with Medication Management and the National Medication List in Sweden: An Interview Study from a Human, Organizational, and Technology Perspective 287
Tora Hammar, Mikael Hoffmann and Lina Nilsson

Local Data Quality Assessments on EHR-Based Real-World Data for Rare Diseases 292
Kais Tahar, Raphael Verbuecheln, Tamara Martin, Holm Graessner and Dagmar Krefting

Agent Based Modelling for Simulating the Interregional Patient Mobility in Italy 297
Fabrizio Pecoraro, Filippo Accordino, Federico Cecconi and Mario Paolucci

Towards a Consistent Representation of Contradictions Within Health Data for Efficient Implementation of Data Quality Assessments 302
Khalid O. Yusuf, Sabine Hanss and Dagmar Krefting

Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned 307
Hans-Ulrich Prokosch, Marie Gebhardt, Julian Gruendner, Philip Kleinert, Karoline Buckow, Lorenz Rosenau and Sebastian C. Semler

A Model for Multi-Institutional Clinical Data Repository 312
Karthik Natarajan, Chunhua Weng and Soumitra Sengupta

Enhancing Data Protection via Auditable Informational Separation of Powers Between Workflow Engine Based Agents: Conceptualization, Implementation, and First Cross-Institutional Experiences 317
Felix Erdfelder, Henning Begerau, David Meyers, Klaus-Jürgen Quast, Daniel Schumacher, Tobias Brieden, Roland Ihle, Danny Ammon, Henner M. Kruse and Sven Zenker

Data Quality Assessment for Observational Medical Outcomes Partnership Common Data Model of Multi-Center 322
Seol Whan Oh, Soo Jeong Ko, Yun Seon Im, Surin Jung, Bo Yeon Choi, Jae Yoon Kim, Wona Choi and In Young Choi

Timeline of and Expectations for the National Medication List in Sweden 327
Mikael Hoffmann

Enriching Remote Monitoring and Care Platforms with Personalized Recommendations to Enhance Gamification and Coaching 332
Antonios Pardos, Parisis Gallos, Andreas Menychtas, Christos Panagopoulos and Ilias Maglogiannis

MedSecurance Project: Advanced Security-for-Safety Assurance for Medical Device IoT (IoMT) 337
Parisis Gallos, Rance DeLong, Nicholas Matragkas, Allan Blanchard, Chokri Mraidha, Gregory Epiphaniou, Carsten Maple, Konstantinos Katzis, Jaime Delgado, Silvia Llorente, Pedro Maló, Bruno Almeida, Andreas Menychtas, Christos Panagopoulos, Ilias Maglogiannis, Petros Papachristou, Mariana Soares, Paula Breia, Ana Cristina Vidal, Martin Ratz, Ross Williamson, Eduard Erwee, Lukasz Stasiak, Orfeu Flores, Carla Clemente, John Mantas, Patrick Weber, Theodoros N. Arvanitis and Scott Hansen

Leveraging Clinical Data Warehouses to Measure Impact of Update Prescription Guidelines of Polyvalent Immunoglobulins in 2018 in France 342
Morgane Pierre-Jean, Guillaume Bouzille, Mathilde Bories, Pascal Le Corre and Marc Cuggia

Electronic Health Records as Information Source in Assessment of the Effectiveness of Delivered Care – A Pilot Study 344
Hanna Von Gerich, Hans Moen and Laura-Maria Peltonen

What Are We Talking About When We Talk About Information-Driven Care? A Delphi-Study on a Definition 346
Elin Siira, Petra Svedberg, Carl Savage and Jens Nygren

Identifying and Predicting Postoperative Infections Based on Readily Available Electronic Health Record Data 348
Siri Lise van der Meijden, Anna van Boekel, Laurens Schinkelshoek, Harry van Goor, Mark de Boer, Ewout Steyerberg, Bart Geerts and Sesmu Arbous

Improving Pressure Ulcers Prediction in Nursing Homes with ML Algorithm 350
Clara Charon, Pierre-Henri Wuillemin and Joël Belmin

Domain Knowledge-Driven Generation of Synthetic Healthcare Data 352
Atiye Sadat Hashemi, Amira Soliman, Jens Lundström and Kobra Etminani

Learning from Health Professionals: A User-Centred Approach to Design a Wound Monitoring Platform 354
Beatriz Félix, Ricardo Melo, Raquel Marques, Paulo Alves and Pedro Faria Salgado

Patients' Experiences of Unwanted Access to Their Online Health Records 356
Annika Bärkås, Anna Kharko, Rose-Mharie Åhlfeldt and Maria Hägglund

RD-MON – Building a Rare Disease Monitor to Enhance Awareness for Patients with Rare Diseases in Intensive Care 358
Romina Blasini, Achim Michel-Backofen, Henning Schneider and Kurt Marquardt

Pharmacists' Experiences of Using Knowledge Bases on Medicines in Pregnancy and Breastfeeding 360
Linnéa Karlsson Lind, Patrik Dreher Sköld and Ulrika Nörby

Local Approval Processes in a Federated and Distributed Research Infrastructure – Lessons Learned from the AKTIN-Project 362
Alexander Kombeiz, Jonas Bienzeisler, Saskia Ehrentreich, Rainer Röhrig, Raphael W. Majeed and Aktin Research Group

Deep Learning Method for Estimation of Morphological Parameters Based on CT Scans 364
Rajarajeswari Ganesan, Antonino A. La Mattina, Frans N. van de Vosse and Wouter Huberts

Towards an Advanced Digital Infrastructure Within the Non-University Sector Demonstrated by the PICOS App 366
Denise Molinnus, Angelique Kurth, Volker Lowitsch, Gernot Marx and Johannes Bickenbach

odML-Tables as a Metadata Standard in Microneurography 368
Alina Troglio, Fabian Schlebusch, Rainer Röhrig, James Dunham, Barbara Namer and Ekaterina Kutafina

Medical Apps for Android and iOS: Differences and Similarities 370
Evgenii Pustozerov, Ute Von Jan, Dennis Lawin and Urs-Vito Albrecht

Implementation of Interoperable Healthcare Standards for Community Healthcare 372
Ye Seul Bae, Yeju Park, Seung Min Lee, Hahyun You, Yujin Park, Hyeonji Lee and Hyung-Jin Yoon

Status of Phlebitis in South Korean Hospitals: Focusing on Electronic Incident Reporting Systems 374
Insook Cho, Jae-Ho Lee and Eun Man Kim

Towards the Definition of an Intelligent System for Organizing Medical Visits and Collecting Medical Data 376
Evangelos Logaras, Antonis Billis, Evangelos Stamkopoulos, Paraskevas Lagakis, Ilias Dimitriadis, Athena Vakali and Panagiotis D. Bamidis

A Framework for Evaluating Synthetic Electronic Health Records 378
Emmanuella Budu, Amira Soliman, Kobra Etminani and Thorsteinn Rögnvaldsson

Protocol of Validation Mechanisms Within the Design of an Electronic Case Report Form for the MULTI-SITA Project 380
Gabriele Di Meco, Daniele Roberto Giacobbe, Sara Mora, Silvia Dettori, Matteo Bassetti, Mauro Giacomini and SITA (Italian Society of Anti-Infective Therapy)

Validation Rules as a First Step for Data Quality: Pharmacovigilance Application in Portugal 382
Bárbara Costa, Bruno Figueiredo and Vera Pinheiro

A Platform Promoting Inter-Physician Interaction to Support the Management of Adverse Drug Reactions for CLL Patients 384
Panos Bonotis and Pantelis Natsiavas

From FAIR4Health Project to 1+MG Initiative: A Spain – Italy Collaboration 386
Ariam Boaglio, Carlos Luis Parra-Calderón, Alicia Martinez-García, María José Escalona-Cuaresma, Norbert Maggi, Domenico Coviello, Paolo Uva and Mauro Giacomini

Barriers, Priorities and Lessons Learned in Achieving Electronic Health Records Interoperability in Low- and Middle-Income Countries: Workshop Findings 388
Taiwo Adedeji, Haythem Nakkas, Hamish Fraser and Philip Scott

Harvesting the Low Hanging Fruits From the FAIRtree 390
Esther Thea Inau, Atinkut Zeleke and Dagmar Waltemath

Text Extraction and Standardization System Development for Pathological Records in the Korea Biobank Network 392
SooJeong Ko, Sunghyeon Park, SeolWhan Oh, YunSeon Im, Surin Jung, BoYeon Choi, Jaeyoon Kim, Wona Choi and InYoung Choi

Refinement of the Target Population for Colorectal Cancer Screening in France, Inventory as a Prelude to the Use of Medico-Administrative Database 394
Akoi Koïvogui, Robert Benamouzig and Catherine Duclos

Transfer Learning for Early Prediction of Adverse Drug Reactions: Docetaxel and Alopecia in Breast Cancer as a Case Study 396
Stella Dimitsaki, Pantelis Natsiavas and Marie-Christine Jaulent

An Audit of Dietitians' Documentation – Comparing the Level of Agreement Between the Audit Instruments Diet-NCP-Audit and NCP-QUEST 398
Anja Lindberg, Maria Nyberg, Ylva Orrevall and Evelina Liljeberg

Section 3. Citizen Health Informatics, Societal Aspects and Education

What Do Autistic People Discuss on Twitter? An Approach Using BERTopic Modelling 403
Elia Gabarron, Enrique Dorronzoro, Daniel Reichenpfader and Kerstin Denecke

Scale Up Multilingualism in Health Emergency Learning: Developing an Automated Transcription and Translation Tool 408
Heini Utunen, Thomas Staubitz, Richelle George, Yu (Ursula) Zhao, Sebastian Serth and Anna Tokar

Extended Reality Is Underutilized in Medical Device Training: A Descriptive Literature Review 413
Minna Laukkavirta, Krista Tarkkala, Marjo Tienari and Tuija S. Ikonen

Informing Family About Patient Trajectory During Surgery: Design and Preliminary Evaluation 418
Frederic Ehrler and Alexandre Jeleff

Evidence of Digital Health Applications from a State-Regulated Repository for Reimbursable Health Applications in Germany 423
Urs-Vito Albrecht, Ute Von Jan, Dennis Lawin, Evgenii Pustozerov and Florian Dittrich

The Inverse Data Law: Market Imperatives, Data, and Quality in AI Supported Care 428
Saleem Ameen, Ming Chao Wong, Kwang Chien Yee, Christian Nøhr and Paul Turner

Streamlining Tangible 3D Printed and Intangible XR Content Creation and Evaluation: The ENTICE Experience 433
Panagiotis Antoniou, Efstathios Sidiropoulos, Natalia Stathakarou, Charalampos Chatzimallis, Evangelos Chondrokostas, Dimas Sumunar, Tobias Karlsson, Sofia Lachanoudi, Panagiotis David, Konstantinos Tagaras, Annita Varella, Alkinoos Athanasiou, James Pickering, Andreas Tooulias, Ion Karolos, Panagiotis-Marios Filippidis, Eirini Schiza, Vassilis Voulgarakis, Charalambos Bratsas, Vassilis Tsioukas, George Tsoulfas and Panagiotis Bamidis

On the Effective Dissemination and Use of Learning Objectives Catalogs for Health Information Curricula Development 438
Oliver J. Bott, Ursula Berger, Nicole Egbert, Carolin Herrmann, Birgit Schneider, Björn Sellemann, Cord Spreckelsen, Brigitte Strahwald, Julian Varghese and Alfred Winter

Use of Digital Games for Educational Purposes Among Medical and Paramedical Sciences Students, Mashhad, Iran 443
Masoumeh Sarbaz, Alireza Banaye Yazdipour, Seyyedeh Fatemeh Mousavi Baigi, Ali Darroudi and Khalil Kimiafar

Perceptions of Learning Activities in Electronic Health Record Transition 448
Thomas Schmidt and Christian Nøhr

Factors to Consider when Introducing Digital Social Activities to Older Persons with Home Care 453
Madeleine Blusi, Helena Lindgren and Ingeborg Nilsson

Feasibility of a Virtual Reality App to Promote Pulmonary Rehabilitation 458
Joseph Finkelstein, Irena Parvanova and Xingye Huo

A Web-Based Public Health Intervention for Addressing Vaccine Misinformation: Analysis of Learner Engagement and Shift in Hesitancy to Vaccinate 463
Leigh Powell, Radwa Nour, Hanan Al Suwaidi and Nabil Zary

Danish Citizens' Expectations for mHealth Prescription 468
Victor Vadmand Jensen, Tom Høg Sørensen, Alice Ørsted Kristensen and Jeppe Eriksen

First Contact, First Learnings – Nursing Staff Approaching Robotics in Health Care 470
Jörn Krückeberg, Daniel Peter Beume, Ronny Klawunn, Jens Müller, Leif Griga, Lena Winkelhake, Lukas Eich, Nina Kempe and Esma Albayrak

Access to Development Opportunities in Biomedical and Health Informatics 472
Sidsel Villumsen, Rikke Amalie Agergaard Jensen, Klaus Nielsen, Ouafa Rian and Charlotte Jonasson

Automated Monitoring Reports of the Activity of the French National Professional Suicide Prevention Helpline 474
Antoine Lamer, Chloé Saint-Dizier, Emile Fares, Christophe Debien, Elise Cleva, Marielle Whatelet and Charles-Edouard Notredame

Health Informatics Training Program in Low and Middle Income Countries 476
Armine Lulejian, Lynn Kysh, Tamara Galoyan and James Dickhoner

Exploring Obese Adults' Preferences for a Physical Activity Chatbot: Qualitative Study 478
Dillys Larbi, Rolf Wynn, Eirik Årsand, Kerstin Denecke, Paolo Zanaboni and Elia Gabarron

N-of-1 Analytics Makerspace: A Prototype for Cystic Fibrosis Self-Care and Co-Care 480
Andreas Hager and Carolina Wannheden

Medical Semiology Teaching Based on Intelligent eLearning 482
Eustache Muteba Ayumba

Support of Young Talent in the GMDS 484
Christina Schüttler, Hannes Ulrich and GMDS Presidential Commission for the Support of Young Scientists

Reduction in X-ray Retake Rate Using the Token Economy Method 486
Hiroto Hada, Shusuke Hiragi, Yukiko Mori, Kenichi Saito, Masumi Ai and Tomohiro Kuroda

Fairness in Artificial Intelligence: Regulatory Sanbox Evaluation of Bias Prevention for ECG Classification 488
Arian Ranjbar, Kristin Skolt, Kathinka Theodore Aakenes Vik, Beate Sletvold Øistad, Eilin Wermundsen Mork and Jesper Ravn

Quantitative Study of a Regional Patient Portal Usage in the Pandemic Period 490
Javier Quiles, Ricardo López and Benigno Rosón

Mobile Application for Improvement of Self-Management of Type 2 Diabetes: Usability Pilot Test 492
Andreia Pinto, João Viana, Glória Conceição, Paulo Santos, Cristina Santos and Alberto Freitas

The Demonstration of a Tool for Self-Estimating Digital Competence 494
Karin Ahlin

E-learning as Part of Residency Education 496
Adam Abovarda, Helena Vallo Hult, Christian Master Östlund and Paul Pålsson

Proposing a Novel Hybrid Short-Term Exchange Program in Biomedical and Health Informatics Education 498
Thomas Engelsma, Carolin Anders, Susan J. Oudbier, Karen Eilbeck, Petra Knaup, Linda W. Peute and Matthias Ganzinger

Expert Feedback on the Adaptation and Translation of Spanish Version of WiseApp 500
Rebecca Schnall, Sergio Ozoria Ramirez, Janeth Juarez Padilla, Mina Halpern, Susan Olender and Pamela Baez

Publication Dynamics on Social Media During the Orpea Nursing Homes Scandal: A Twitter Analysis 502
Mathilde Fruchart, Laetitia Verdier, Jean-Baptiste Beuscart and Antoine Lamer

Digital Skills Among Elderly Care Workers 504
Eija Kivekäs, Ulla-Mari Kinnunen, Johanna Ikonen and Kaija Saranto

The Complexity of a Clinical History 506
Antonio D'Uffizi, Fabrizio Pecoraro, Fabrizio L. Ricci, Giuseppe Stecca and Fabrizio Consorti

Section 4. Decision Support Systems

The Effect of Physicians' Acknowledgement of Clinical Decision Support Systems Generated Alerts on Patient Diabetes Management in a Primary Care Setting 511
Mohammad A. Faysel, Tameca Miller, Jonathan Singer and Caroline Cummings

Secur-e-Health Project: Towards Federated Learning for Smart Pediatric Care 516
Rita Rb-Silva, Xavier Ribeiro, Francisca Almeida, Carolina Ameijeiras-Rodriguez, Julio Souza, Luis Conceição, Tiago Taveira-Gomes, Goreti Marreiros and Alberto Freitas

PDSS: A Pharmacological Decision Support System for Diabetics Patients with COVID-19 521
Isabel Amaya-Rodriguez, Nekane Larburu, María Rollán Martinez-Herrera, Kristin Rebescher, Iván Macia, Miguel A. Armengol De La Hoz, Cristina Rubio-Escudero and Alba Garin-Muga

Parralel Recurrent Convolutional Neural Network for Abnormal Heart Sound Classification 526
Arash Gharehbaghi, Elaheh Partovi and Ankica Babic

Clinical Decision Support: Evaluating the Development of a Tool for Nurses 531
Bente Christensen and Ann Kristin Rotegård

Prediction of COVID-19 Mortality in the Intensive Care Unit Using Machine Learning 536
Aikaterini Sakagianni, Christina Koufopoulou, Vassilios Verykios, Evangelos Loupelis, Dimitrios Kalles and Georgios Feretzakis

Interdisciplinary Teams in Health Informatics: Using FHIR Standards to Share Computable Knowledge 541
Elisavet Andrikopoulou, Björn Schreiweis and Michael Anywar

Extracting Temporal Relationships in EHR: Application to COVID-19 Patients 546
Carlos Molina and Belén Prados-Suarez

The PrescIT Knowledge Graph: Supporting ePrescription to Prevent Adverse Drug Reactions 551
Achilleas Chytas, Vlasios Dimitriadis, Giorgos Giannios, Margarita Grammatikopoulou, George Nikolaidis, Jenny Pliatsika, Martha Zachariadou, Haralampos Karanikas, Ioannis Kompatsiaris, Spiros Nikolopoulos and Pantelis Natsiavas

Interdisciplinary Human-Centered AI for Hospital Readmission Prediction of Heart Failure Patients 556
Amira Soliman, Monika Nair, Marcus Petersson, Lina Lundgren, Petra Dryselius, Ebba Fogelberg, Omar Hamed, Kobra Etminani and Jens Nygren

Patient Electronic Health Record as Temporal Graphs for Health Monitoring 561
Hugo Le Baher, Jérôme Azé, Sandra Bringay, Pascal Poncelet, Nancy Rodriguez and Caroline Dunoyer

Feature Selection Based on a Genetic Algorithm for Optimizing Weaning Success 566
Samanta Rosati, Andrea Scotto, Vito Fanelli and Gabriella Balestra

Towards an Explainable AI-Based Tool to Predict Preterm Birth 571
Ilias Kyparissidis Kokkinidis, Evangelos Logaras, Emmanouil S. Rigas, Ioannis Tsakiridis, Themistoklis Dagklis, Antonis Billis and Panagiotis D. Bamidis

Convolutional Neural Networks for Optical Discrimination Between Histological Types of Colorectal Polyps Based on White Light Endoscopic Images 576
Vasileios Panteris, Georgios Feretzakis, Panagiotis Karantanos, Dimitris Kalles, Vassilios V. Verykios, Maria Panoutsakou, Eirini Karagianni, Christina Zoubouli, Stefani Vgenopoulou, Aikaterini Pierrakou, Maria Theodorakopoulou, Apostolos E. Papalois, Thomas Thomaidis, Ilias Dalainas and Elias Kouroumalis

Blood Vessel Segmentation Using U-Net for Glaucoma Diagnosis with Limited Data 581
Lukas Schiesser, Jens Julian Storp, Kemal Yildirim, Julian Varghese and Nicole Eter

First Steps Towards a Risk of Bias Corpus of Randomized Controlled Trials 586
Anjani Dhrangadhariya, Roger Hilfiker, Martin Sattelmayer, Katia Giacomino, Rahel Caliesch, Simone Elsig, Nona Naderi and Henning Müller

Developing a Comprehensive Search Strategy for the Systematic Review of Clinical Decision Support Systems for Nursing Practice 591
Cynthia Abi Khalil, Antoine Saab, Jihane Rahme and Brigitte Seroussi

Qualitative Assessment of Implementation of a Discharge Prediction Tool Using RE-AIM Framework 596
Joseph Finkelstein, Irena Parvanova, Zhaopeng Xing, Tuyet-Trinh Truong and Andrew Dunn

Decision Support for Signal Assessment of Large Case Series in Pharmacovigilance 601
Lucy Quirant, Lovisa Sandberg, Jim W. Barrett and Johan Ellenius

Explainable Graph Neural Networks for Atherosclerotic Cardiovascular Disease 603
Jens Lundström, Atiye Sadat Hashemi and Prayag Tiwari

Guideline-Based Algorithmic Recommendations Versus Multidisciplinary Team Advice for Gynecologic Oncology 605
Kees Ebben, Marleen van Houdt, Cor de Kroon and Jurrian van der Werf

Diagnosis Support for Rare Diseases Using Phenotypic Profiles 607
Richard Noll, Jannik Schaaf, Michael Von Wagner and Holger Storf

A Masked Language Model for Multi-Source EHR Trajectories Contextual Representation Learning 609
Ali Amirahmadi, Mattias Ohlsson, Kobra Etminani, Olle Melander and Jonas Björk

Modeling Clinical Guidelines for an Epilepsy-CDSS: The EDiTh Project 611
Ariadna Pérez Garriga, Stefan Wolking, Jonas Fortmann, Raphael W. Majeed, Christian Stockem, Lukas Niekrenz, Christian Bosselmann, Yvonne Weber, Rainer Röhrig and Myriam Lipprandt

Temporal Context Matters: An Explainable Model for Medical Resource Utilization in Chronic Kidney Disease 613
Omar Hamed, Amira Soliman and Kobra Etminani

An Integrated Approach to Automated Diagnosis of Cervical Intraepithelial Neoplasia in Digital Histology Images 615
Bríd Brosnan, Inna Skarga-Bandurova, Tetiana Biloborodova and Illia Skarha-Bandurov

Digital Information Management for Advanced Practice Nursing: Needs Assessment 617
Heidi Laine, Asta Heikkilä, Laura-Maria Peltonen and Jouni Smed

Section 5. Human Factors and Organisational Issues

Me-to-We Design: A Blueprint for Enriching Welfare Technologies 621
Morten Hertzum and Mika Yasuoka

Usability Engineering of Dynamic Biosignal Displays Using Ventilation Data 626
Beatrice Coldewey, Anne Marie Klöckener, Christof Göbel, Rainer Röhrig and Myriam Lipprandt

Eye-Tracking on Touch Screen – Evaluating User Interaction 631
Camilla Panduro Nielsen, Jacob Nielsen, Chris Bang Sørensen and Ellen Raben Pedersen

Inter-Professional Communications During Follow-Up of Type 2 Diabetes Patients: An Exploratory Study 636
Romaric Marcilly, Madleen Lemaitre, Sonia Maalem, Jan Baran, Jean-Baptiste Beuscart and Paul Quindroit

How to Design Successful Participatory Design Workshops for Digital Health Solutions? 641
Kerstin Denecke, François von Kaenel, Marko Miletic, Carlos Fernández-Llatas, Gema Ibañez-Sánchez, Zoe Valero-Ramón, Antonio Martînez-Millana, María Segura and Octavio Rivera Romero

Enhancing Inclusive mHealth Design for People Living with Dementia: Examples from Literature 646
Thomas Engelsma, Monique Jaspers and Linda Peute

Using Real-Time Interaction Analysis to Explore Human-Robot Interaction 651
Jiye Kang, Su Jin Kim, Su Hyun Moon, Su Min Kim, Yujin Seo, Won Chul Cha and Meong Hi Son

Methodology for the Description of Socio-Technical Systems: A Case Study Approach 656
Myriam Lipprandt, Andrea D. Klausen, Rainer Röhrig and DESIREE Study Group

How Do User Participation and IT Self-Efficacy Influence User Attitudes Towards Smart Hospital Technology? 661
Saskia Kröner, Jörg Hassmann, Moritz Esdar, Julia Maischak and Ursula Hübner

Improving Well-Being in Schools – Lessons Learned from IoT Experts 666
Agnieszka Kitkowska and Karin Ahlin

How Do Informaticians and IT-Architects Collaborate, or Not? A Case Study from a Public Health Care Provider 671
Anna Rossander

Ethical Perspectives on Implementing AI to Predict Mortality Risk in Emergency Department Patients: A Qualitative Study 676
Lena Petersson, Kalista Vincent, Petra Svedberg, Jens M. Nygren and Ingrid Larsson

Healthcare Leaders' Perceptions of the Usefulness of AI Applications in Clinical Work: A Qualitative Study 678
Lena Petersson, Petra Svedberg, Jens M. Nygren and Ingrid Larsson

Usability-Tests of Mechanical Ventilators: A Systematic Review 680
Julia S. Volmerg, Beatrice Coldewey, Rainer Röhrig and Myriam Lipprandt

Towards Implementation of a Home-Based Phantom Limb Pain Treatment Facilitated by Textile-Electrode System – A Case Study 682
Anna Björkquist, Li Guo, Morten B. Kristoffersen, Maria M. Novoa, Max Ortiz-Catalan and Leif Sandsjö

Usability Assessment as a Participatory Method for Implementing Technical Innovations in Nursing Care 684
Jörn Krückeberg, Isabel Ottmann, Daniel Peter Beume, Ronny Klawunn and Nicole Hechtel

Physician-Reported Experience and Usability of the MyPal Platform: A Palliative Care Digital Health Intervention 686
Panos Bonotis and Pantelis Natsiavas

Section 6. Knowledge and Information Representation and Modeling

Machine Learning for Medical Data Integration 691
Armin Müller, Lara-Sophie Christmann, Severin Kohler, Roland Eils and Fabian Prasser

Definition, Composition, and Harmonization of Core Datasets Within the German Center for Lung Research 696
Mark R. Stöhr, Andreas Günther and Raphael W. Majeed

Building a Disease Knowledge Graph 701
Enayat Rajabi and Somayeh Kafaie

Performance Benchmarking of FHIR Terminology Operations in ETL Jobs 706
Joshua Wiedekopf, Cora Drenkhahn and Josef Ingenerf

Fitness for Use of Anatomical Therapeutic Chemical Classification for Real World Data Research 711
Ines Reinecke, Elisa Henke, Yuan Peng, Martin Sedlmayr and Franziska Bathelt

Observational Cohort Study Dedicated to Autism Spectrum Disorder: Milestone Steps, Results Updates, Perspectives 716
Mohamed Ben Said, Laurence Robel, Fanny Kpenou, Jean-Philippe Jais, Mario Speranza and TEDIS Study Group

Concept for a Basic ISO 14721 Archive Information Package for Clinical Studies 721
Matthias Löbe, Alessandra Kuntz, Christian Henke, Frank Meinke, Ulrich Sax and Alfred Winter

Model and Strategy for Predicting and Discovering Drug-Drug Interactions 726
Abdelmalek Mouazer, Nada Boudegzdame, Karima Sedki, Rosy Tsopra and Jean-Baptiste Lamy

Building an Ontology for Traditional Medicine by Comparing Traditional Medicine Information of Chapter 26 of ICD-11 with SNOMED CT 731
Tom Lundholm and Stefano Bonacina

Development of Verified Innovation Process for Healthcare Solutions (VIPHS): A Stepwise Model for Digital Health 736
Eunji Lee, Bengt Arne Sjöqvist, Magnus Andersson Hagiwara, Hanna Maurin Söderholm and Stefan Candefjord

Insights into the FAIRness of the German Network University Medicine: A Survey 741
Lea Michaelis, Rasim Atakan Poyraz, Michael Rusgoza Muzoora, Kerstin Gierend, Alexander Bartschke, Christoph Dieterich, Tim Johann, Dagmar Krefting, Dagmar Waltemath and Sylvia Thun

Health-Related Content in Transformer-Based Language Models: Exploring Bias in Domain General vs. Domain Specific Training Sets 743
Giuseppe Samo and Caterina Bonan

Ontological Modeling of Clinical Study Forms 745
Jacques Hilbey, Amel Raboudi, Marie-Odile Krebs and Jean Charlet

Exploring Use Cases for CovidGraph 747
Lea Gütebier, Ron Henkel, Christian Niklas and Dagmar Waltemath

Exploring New Possibilities for Research Data Exploration Using the Example of the German Core Data Set 749
Felix Menzel, Dagmar Waltemath and Ron Henkel

Towards ETL Processes to OMOP CDM Using Metadata and Modularization 751
Yuan Peng, Elisa Henke, Martin Sedlmayr and Franziska Bathelt

Evaluation of Modeling Approaches for a Clinical Data Warehouse in a Highly Dynamic Environment 753
Stephan Lorenz, Richard Gebler, Franziska Bathelt, Martin Sedlmayr and Ines Reinecke

Modeling Patient Treatment Trajectories Using Markov Chains for Cost Analysis 755
Markus Haug, Raivo Kolde, Marek Oja and Maarja Pajusalu

Representing Sex and Gender Information in Biomedical Research 757
Dagmar Waltemath, Hilke Beelich, Philipp Töpfer, Corinna Bergelt, Agnes Flöel, Elpiniki Katsari, Elke Krüger, Stine Lutze, Michael Schöner, Sebastian Paschen, Moritz Roloff, Hanne Roth and Sylvia Stracke

A Categorial Structure for Identifying Physiological Measurement Observables 759
Anders Thurin

Section 7. Natural Language Processing

Automated ICF Coding of Rehabilitation Notes for Low-Resource Languages via Continual Training of Language Models 763
Kevin Roitero, Andrea Martinuzzi, Maria Teresa Armellin, Gabriella Paparella, Alberto Maniero and Vincenzo Della Mea

Evaluating the Portability of Rheumatoid Arthritis Phenotyping Algorithms: A Case Study on French EHRs 768
Thibaut Fabacher, Erik-André Sauleau, Noémie Leclerc Du Sablon, Hugo Bergier, Jacques-Eric Gottenberg, Adrien Coulet and Aurélie Névéol

Learning to Classify Medical Discharge Summaries According to ICD-9 773
Leonardo Moros, Jérôme Azé, Sandra Bringay, Pascal Poncelet, Maximilien Servajean and Caroline Dunoyer

User-Centered Design of a Speech-Based Application to Support Caregivers 778
Daniel Reichenpfader and Matthias Gira

Investigating Canadian Public Attitudes Toward COVID-19 Vaccine Mandates with a Nested Analysis Framework 783
Yang Yang, Xiaomeng Lei, Zeping Zhou, Shu-Feng Tsao, Zahid A. Butt and Helen H. Chen

Secondary Use of Clinical Problem List Entries for Neural Network-Based Disease Code Assignment 788
Markus Kreuzthaler, Bastian Pfeifer, Diether Kramer and Stefan Schulz

Ontology-Based Semantic Annotation of French Psychiatric Clinical Documents 793
Ons Aouina, Jacques Hilbey and Jean Charlet

The Prevalence of mRNA Related Discussions During the Post-COVID-19 Era 798
Dimitrios Kokkinakis, Bastiaan Bruinsma and Mia-Marie Hammarlin

Classification of Clinical Notes from a Heart Failure Telehealth Network 803
Fabian Wiesmüller, Aaron Lauschenski, Martin Baumgartner, Dieter Hayn, Karl Kreiner, Bettina Fetz, Luca Brunelli, Gerhard Pölzl, Bernhard Pfeifer, Sabrina Neururer and Günter Schreier

Classifiers of Medical Eponymy in Scientific Texts 808
Dennis Toddenroth

Communicating in Emergency Settings: BabelDr or Telephone Interpreting? 813
Céline De Graaf, Pierrette Bouillon and Hervé Spechbach

Diagnosis Classification in the Emergency Room Using Natural Language Processing 815
Marieke M. van Buchem, Hanna H. 't Hart, Pablo J. Mosteiro, Ilse M.J. Kant and Martijn P. Bauer

Predicting Depression Risk in Patients with Cancer Using Multimodal Data 817
Anne de Hond, Marieke van Buchem, Claudio Fanconi, Mohana Roy, Douglas Blayney, Ilse Kant, Ewout Steyerberg and Tina Hernandez-Boussard

Few-Shot and Prompt Training for Text Classification in German Doctor's Letters 819
Phillip Richter-Pechanski, Philipp Wiesenbach, Dominic M. Schwab, Christina Kiriakou, Mingyang He, Nicolas A. Geis, Anette Frank and Christoph Dieterich

In-Hospital Cancer Mortality Prediction by Multimodal Learning of Non-English Clinical Texts 821
Shintaro Oyama, Taiki Furukawa, Shotaro Misawa, Ryuji Kano, Hirokazu Yarimizu, Tomoki Taniguchi, Kohei Onoda, Kikue Sato and Yoshimune Shiratori

Translating Medical Dialogues into Pictographs: An Approach Using UMLS 823
Johanna Gerlach, Pierrette Bouillon, Magali Norré and Hervé Spechbach

SapBERT-Based Medical Concept Normalization Using SNOMED CT 825
Akhila Abdulnazar, Markus Kreuzthaler, Roland Roller and Stefan Schulz

Data-Driven Identification of Clinical Real-World Expressions Linked to ICD 827
Amila Kugic, Bastian Pfeifer, Stefan Schulz and Markus Kreuzthaler

Digital Health Data Capture with a Controlled Natural Language 829
Kristian Juha Ismo Kankainen, Inari Listenmaa, Gunnar Piho and Peeter Ross

Information Extraction from Medical Texts with BERT Using Human-in-the-Loop Labeling 831
Hendrik Šuvalov, Sven Laur and Raivo Kolde

Large Language Model as Unsupervised Health Information Retriever 833
Keyuan Jiang, Mohammed M. Mujtaba and Gordon R. Bernard

Announcement of the German Medical Text Corpus Project (GeMTeX) 835
Frank Meineke, Luise Modersohn, Markus Loeffler and Martin Boeker

Clustering Similar Diagnosis Terms 837
Stefan Schulz, Akhila Abdulnazar and Markus Kreuzthaler

Section 8. Public Health and Epidemiology Informatics

Health Research Requires Efficient Platforms for Data Collection from Personal Devices 841
Erlend Johannessen, André Henriksen, Eirik Årsand, Alexander Horsch, Jonas Johansson and Gunnar Hartvigsen

Standardization Proposal for the Transmission of Waiting List Data in Italy 846
Ylenia Murgia, Monica Bonetto, Roberta Gazzarata, Luisa Brogonzoli, Rosaria Iardino, Alessandro Venturi and Mauro Giacomini

OpenDataPsy: An Open-Data Repository with Standardized Storage and Description for Research in Psychiatry 851
Chloé Saint-Dizier, Antoine Lamer, Majda Zaanouar, Alina Amariei and Paul Quindroit

Description of a French Population of Diabetics Treated Followed up by General Practitioners 856
Mathilde Fruchart, Antoine Lamer, Madleen Lemaitre, Jean-Baptiste Beuscart, Matthieu Calafiore and Paul Quindroit

Digital Disease Surveillance for Emerging Infectious Diseases: An Early Warning System Using the Internet and Social Media Data for COVID-19 Forecasting in Canada 861
Yang Yang, Shu-Feng Tsao, Mohammad A. Basri, Helen H. Chen and Zahid A. Butt

Using Big Data to Uncover Association Between Sildenafil Use and Reduced Risk of Alzheimer's Disease 866
Xingyue Huo and Joseph Finkelstein

Conducting an Epidemiologic Study and Making It FAIR: Reusable Tools and Procedures from a Population-Based Cohort Study 871
Carsten Oliver Schmidt, Stephan Struckmann, Maik Scholz, Janka Schössow, Dörte Radke, Adrian Richter, Achim Reineke, Elisa Kasbohm, Joany Mariño Coronado, Birgit Schauer, Kristin Henselin, Susanne Westphal, Darko Balke, Torsten Leddig, Henry Völzke and Jörg Henke

Data Collection and Analysis Methods for Smart Nudging to Promote Physical Activity: Protocol for a Mixed Methods Study 876
Seshathiri Dhanasekaran, Anders Andersen, Randi Karlsen, Anne Håkansson and André Henriksen

Modelling Information Needs and Sources in a COVID-19 Designated Hospital 881
Elizabeth M. Borycki, Andre W. Kushniruk, Henry Oluka, Simon Minshall, Kenrick Cato, Yalini Senathirajah and David Kaufman

Self, at Home, and Digital if Possible, Drivers to Population Health Management 886
Felix Cillessen and Jacob Hofdijk

An Adaptive Digital Intelligence System to Support Infodemic Management: The WHO EARS Platform 891
Becky White, Ivan Cabalero, Tim Nguyen, Sylvie Briand, Agnese Pastorino and Tina D. Purnat

Narrative Trends over the COVID-19 Pandemic: Digital Social Listening to Inform WHO Infodemic Management 893
Becky White, Atsuyoshi Ishizumi, Brian Yau, Amy Wright, Lucy Lavery, Tim Nguyen, Sylvie Briand, Tim Zecchin and Tina D. Purnat

'Smart' Buffalo Weight Estimation via Digital Technologies: Experiences from South Italy 895
Oscar Tamburis, Roberta Matera, Angela Salzano, Antonio Calamo, Domenico Vistocco and Gianluca Neglia

Identification of Subphenotypes of Opioid Use Disorder Using Unsupervised Machine Learning 897
Fatemeh Shah-Mohammadi and Joseph Finkelstein

Converting HL7 CDA Based Nationwide Austrian Medication Data to OMOP CDM 899
Florian Katsch, Rada Hussein, Raffael Korntheuer and Georg Duftschmid

Prognostic Factors for Covid-19 on Admission Profile and Air Pollutants 901
Kikue Sato, Taiki Furukawa, Satoshi Yamashita, Daisuke Kobayashi, Shintaro Oyama and Yoshimune Shiratori

Prediction of Mental Health Support of Employee Perceiving by Using Machine Learning Methods 903
Hossein Jamalirad and Mahdie Jajroudi

Improving Antibiotic Prescribing for Dentistry in France Using an Ontology 905
Jean-Alexis Buvat, Francesco Monti, Tariq Nsais and Bénédicte Melot

Comparing Responses to COVID-19 Across Institutions: Conceptualization of an Emergency Response Maturity Model 907
Yalini Senathirajah, David Kaufman, Elizabeth Borycki, Andre Kushniruk and Kenrick Cato

Opportunistic Screening for Osteoporosis Using CT Scans of the Knee: A Pilot Study 909
Mahmoud Elmahdy and Ronnie Sebro

Opportunistic Screening for Osteoporosis Using Hand Radiographs: A Preliminary Study 911
Farid Ghareh Mohammadi and Ronnie Sebro

Linkage Health and Environmental Data: A Case Study on Asthma Prevalence in Children and Adolescents in Slovenia 913
Tanja Rejc, Andreja Kukec, Tanja Carli and Ajda Mlakar

Section 9. Telehealth, Sensors, Signals and Imaging Informatics

Towards Automated COVID-19 Presence and Severity Classification 917
Dominik Mueller, Silvan Mertes, Niklas Schroeter, Fabio Hellmann, Miriam Elia, Bernhard Bauer, Wolfgang Reif, Elisabeth André and Frank Kramer

Designing a Persuasive E-Coaching Application for Informal Caregivers 922
Shweta Premanandan, Awais Ahmad, Åsa Cajander and Pär Ågerfalk

Can Synthetic Images Improve CNN Performance in Wound Image Classification? 927
Leila Malihi, Ursula Hübner, Mats L. Richter, Maurice Moelleken, Mareike Przysucha, Dorothee Busch, Jan Heggemann, Guido Hafer, Stefan Wiemeyer, Gunther Heidemann, Joachim Dissemond, Cornelia Erfurt-Berge, Carlotta Barkhau, Achim Hendriks and Jens Hüsers

Few-Shot Meta-Learning for Recognizing Facial Phenotypes of Genetic Disorders 932
Ömer Sümer, Fabio Hellmann, Alexander Hustinx, Tzung-Chien Hsieh, Elisabeth André and Peter Krawitz

Model-Driven Dementia Prevention and Intervention Platform 937
Markus Schinle, Michael Dietrich, Simon Stock, Marius Gerdes and Wilhelm Stork

The Future of Online Video Consultations in Primary Care: A Qualitative Study 942
Maria Hägglund, Nadia Davoody, Anna Kristensson Ekwall and Nasim Farrokhnia

U-Net-Based Segmentation of Current Imaging Biomarkers in OCT-Scans of Patients with Age Related Macular Degeneration 947
Kemal Yildirim, Sami Al-Nawaiseh, Sophia Ehlers, Lukas Schießer, Michael Storck, Tobias Brix, Nicole Eter and Julian Varghese

Random Forest and Gradient Boosted Trees for Patient Individualized Contrast Agent Dose Reduction in CT Angiography 952
René Pallenberg, Marja Fleitmann, Andreas Martin Stroth, Jan Gerlach, Alexander Fürschke, Jörg Barkhausen, Arpad Bischof and Heinz Handels

V-care: An Application to Support Lifestyle Improvement in Children with Obesity 957
Cristiana Larizza, Silvana Quaglini, Michelangelo Chasseur, Valentina Bevolo, Gianvincenzo Zuccotti and Valeria Calcaterra

Reliability of IMU-Derived Gait Parameters in Foot Drop Patients 962
Armando Coccia, Federica Amitrano, Gaetano Pagano, Arcangelo Biancardi, Giuseppe Tombolini and Giovanni D'Addio

On the Difficulty of Predicting Engagement with Digital Health for Substance Use 967
Franziska Günther, Christopher Yau, Sarah Elison-Davies and David Wong

The Assessment of Glioblastoma Metabolic Activity via ^{11}C-Methionine PET and Radiomics 972
Gleb Danilov, Diana Kalaeva, Nina Vikhrova, Tatiana Konakova, Anna Zagorodnova, Angelina Popova, Andrey Postnov, Svetlana Shugay, Michael Shifrin and Igor Pronin

Benchmarking the Impact of Noise on Deep Learning-Based Classification of Atrial Fibrillation in 12-Lead ECG 977
Theresa Bender, Philip Gemke, Ennio Idrobo-Avila, Henning Dathe, Dagmar Krefting and Nicolai Spicher

Patient Perspectives on Long-Term Use of a Pulmonary Telerehabilitation Platform: A Qualitative Analysis 982
Aileen S. Gabriel, Irena Parvanova and Joseph Finkelstein

Telemedicine Based on Human Activity Recognition in Elderly Healthcare 987
Raheleh Ghouchan Nezhad Noor Nia, Afrooz Arzehgar, Vajiheh Dehdeleh and Saeid Eslami

Interpretable EEG-Based Emotion Recognition Using Fuzzy Cognitive Maps 992
Georgia Sovatzidi and Dimitris K. Iakovidis

Client-Side Application of Deep Learning Models Through Teleradiology 997
Sébastien Jodogne

Technological Assessment of Smart Wearables and 5G-Integrated Edge Computing for Real-Time Health Monitoring 1002
Paulo Haas and Thomas M. Deserno

AI-Based Gut-Brain Axis Digital Twins 1007
Stéphane Meystre, Ruud van Stiphout, Annelies Goris and Santiago Gaitan

Associations Between Engagement with the BitHabit Digital Lifestyle Intervention and Changes in Type 2 Diabetes Risk Factors 1009
Hilkka Liedes, Elina Mattila, Anita Honka, Pilvikki Absetz, Kirsikka Aittola, Suvi Manninen, Niina Lintu, Ursula Schwab, Aino-Maija Eloranta, Adil Umer, Tanja Tilles-Tirkkonen, Reija Männikkö, Ilona Ruotsalainen, Mark Van Gils, Jaana Lindström, Timo Lakka, Jussi Pihlajamäki and Anna-Leena Vuorinen

Patients' Introduction to Online Video Consultations in Primary Healthcare 1011
Irene Muli, Marina Taloyan, Lovisa Jäderlund Hagstedt and Maria Hägglund

A Remote Monitoring Platform for the Management of Lower Limb Vascular Diseases 1013
Julio Souza, Ana Vieira, Luís Conceição, Rafael Martins, Daniel Rodrigues, Gustavo Corrente, William Xavier, Sérgio Sampaio, Goreti Marreiros and Alberto Freitas

Trajectories of Anxiety Symptoms for COVID-19 Patients Using Multimodal Data Collected from Telemedicine 1015
Sumi Sung, Su Hwan Kim, Youlim Kim, Mirinae You, Hannah Lee, Soojin Her, Ye Seul Bae and Eui Kyu Chie

Requirements to mHealth Applications for Animal Owners: A Narrative Review 1017
Laura Haase

Implementing Remote Patient Monitoring for Patients During Systemic Cancer Therapy 1019
Lara Kohn, Anita Mueller, Veit Scheble, Philip Storz, Hanna Borlinghaus, Selcan Behiye Ulas, Nisar Malek and Christian Thies

Interest in and Experience with the Use of Patient Portals Among Adolescents in Mental Health Care 1021
Martine Stecher Nielsen, Aslak Steinsbekk and Torunn Hatlen Nøst

Automated Classification of Exercise Exertion Levels Based on Real-Time Wearable Physiological Signal Monitoring 1023
Aref Smiley, Ihor Havrylchuk, Te-Yi Tsai, Elena Zakashansky, Aileen Gabriel, Taulant Xhakli, Wanting Cui, Irena Parvanova, Hu Cui and Joseph Finkelstein

Spectral Fusion of Heartbeat and Accelerometer Data for Estimation of Breathing Rate in Wearable Patches 1025
Mohamed Alhaskir, Jennifer Bauer, Florian Linke, Elisabeth Schriewer, Yvonne Weber, Stefan Wolking, Rainer Röhrig, Markus Rothermel, Henner Koch and Ekaterina Kutafina

Voxel-Wise Medical Imaging Transformation and Adaption Based on CycleGAN and Score-Based Diffusion 1027
Feifei Li, Mirjam Schöneck, Oya Beyan and Liliana Lourenco Caldeira

Effects of Ankle-Foot Orthosis on Balance of Foot Drop Patients 1029
Federica Amitrano, Armando Coccia, Gaetano Pagano, Arcangelo Biancardi, Giuseppe Tombolini and Giovanni D'Addio

On-Site Visualization of Ballistocardiography Data 1031
Urs-Vito Albrecht and Ulf Kulau

Tracking of Nutritional Intake Using Artificial Intelligence 1033
Marko Petković, Joyce Maas and Milan Petković

Section 10. Bioinformatics

Improving Patient Similarity Using Different Modalities of Phenotypes Extracted from Clinical Narratives 1037
Xiaoyi Chen, Carole Faviez, Marc Vincent, Sophie Saunier, Nicolas Garcelon and Anita Burgun

OmicSDK-Transcriptomics: A Web Platform for Transcriptomics Data Analysis 1042
Aurora Sucre, Maria Martinez and Alba Garin-Muga

Differential Gene Expression Data Analysis of ASD Using Random Forest 1047
Pragya, Praveen Kumar Govarthan, Kshitij Sinha, Sudip Mukherjee and Jac Fredo Agastinose Ronickom

Towards a Digital Twin in Human Brain: Brain Tumor Detection Using K-Means 1052
Anastasios Loukas Sarris, Efstathios Sidiropoulos, Evangelos Paraskevopoulos and Panagiotis Bamidis

Challenges of Estimating Global Feature Importance in Real-World Health Care Data 1057
Aniek F. Markus, Egill A. Fridgeirsson, Jan A. Kors, Katia M.C. Verhamme and Peter R. Rijnbeek

How Good Is ChatGPT for Medication Evidence Synthesis? 1062
Hao Liu, Yifan Peng and Chunhua Weng

Combining Sequence Similarity with Physicochemical Properties to Predict Binders for MHC-II Molecules 1067
Ya-Lin Chen

Electromagnetic Fields Literature Analysis for Precision Medicine 1069
Christopher Hawthorne, Donal Harrison, Eva McEvoy and Guillermo Lopez Campos

A Reliable and Secure Method for Sharing Genomic Data 1071
Dinis. B. Cruz, João R. Almeida, Jorge M. Silva and José L. Oliveira

Mapping Exposome Derived Phenotypes into SNOMED Codes 1073
Christopher Hawthorne, Luis Marco Ruiz and Guillermo Lopez Campos

The RCX Extension Hub: A Resource for Implementations Extending the R Adaption of the Cytoscape Exchange Format 1075
Florian Auer, Nadja Bramkamp (Beckmann), Simone Mayer, Dominik Müller and Frank Kramer

Subject Index 1077

Author Index 1089

This page intentionally left blank

Section 1

Caring is Sharing – Exploiting Value in Data for Health and Innovation

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230053

German Claims Data for Real-World Research: Content Coverage Evaluation in OMOP CDM

Elisa HENKE[a,1], Michéle ZOCH[a], Ines REINECKE[a], Melissa SPODEN[b], Thomas RUHNKE[b], Christian GÜNSTER[b], Martin SEDLMAYR[a], and Franziska BATHELT[a]
[a] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*
[b] *AOK Research Institute, WIdO, Berlin, Germany*

Abstract. Research on real-world data is becoming increasingly important. The current restriction to clinical data in Germany limits the view of the patient. To gain comprehensive insights, claims data can be added to the existing knowledge. However, standardized transfer of German claims data into OMOP CDM is currently not possible. In this paper, we conducted an evaluation regarding the coverage of source vocabularies and data elements of German claims data in OMOP CDM. We point out the need to extend vocabularies and mappings to support research on German claims data.

Keywords. OMOP CDM, OHDSI, interoperability, claims data

1. Introduction

Research based on real-world data is becoming increasingly important to gain new insights for personalized diagnoses and treatments. In this context, the German Federal Ministry of Education and Research (BMBF) has extensively funded the Medical Informatics Initiative [1] to provide digital infrastructures for the integration and harmonization of health data. This is currently limited to university hospitals and their patient data. However, this limits the view of patients to the time of hospitalization. This only relates to a small percentage of patients. The far greater number of patients and treatments take place outside the hospital [2]. To get a comprehensive view of the patient's journey, clinical data must be combined with outpatient data. An important data source for outpatient data are health insurance claims data. Although claims data and clinical data differ in structure and periodicity, they can be linked using the patients' unique health insurance number. To provide a complete picture of a patient, the two different datasets have to be merged into a common data model, such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)

[1] Corresponding Author: Elisa Henke, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany; E-mail: Elisa.Henke@tu-dresden.de.

[3,4]. OMOP CDM comes with internationally standardized terminologies and a wide range of tools that enable statistical analyzes and machine learning. While many ways exist to import clinical data into OMOP CDM [5,6], there is no way yet for German claims data. However, the conformance is necessary for using the available analysis tools and the international comparability. To gain initial insights into the current content coverage of German claims data in OMOP CDM, we conducted an evaluation of the extent to which source vocabularies and data elements can be mapped to OMOP CDM.

2. Methods

To ensure successful harmonization of German claims data in OMOP CDM, we focused the two central components of OMOP CDM: the OHDSI standardized vocabularies and the data tables. Transforming German claims data to OMOP CDM requires the standardization of national vocabularies and the mapping of source data elements into OMOP CDM data tables. To examine the current content coverage of German claims data vocabularies and data elements in OMOP CDM, we focused on the following preliminary considerations (Figure 1). We 1) first performed a data profiling to get an overview of German claims data, especially their structure, content and references between data elements. Then 2) we compared German claims data and clinical data to check similarities that allow reuse of existing mappings. Finally, 3) a feasibility analysis of mapping German claims data to OMOP CDM was done to identify any current obstacles that prevent the mapping.

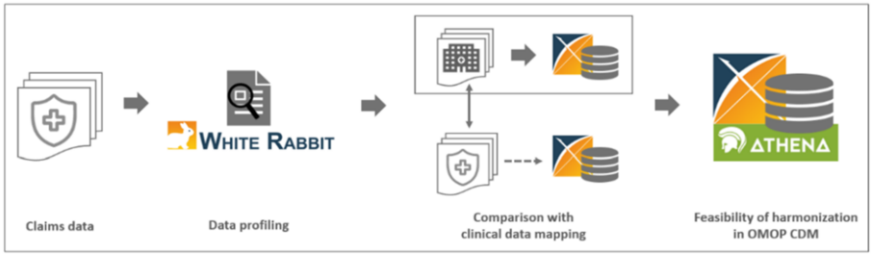


Figure 1. Concept on content coverage evaluation of German claims data in OMOP CDM (own illustration)

2.1. Data profiling

In Germany, healthcare providers and institutes are legally bound to transfer claims data to health insurance companies for billing purposes using a national standardized format. To gain a detailed insight into German claims data, we used synthetic claims data from the German local health care funds (Allgemeine Ortskrankenkassen, AOK), comprising data from 10.000 patients. This data set included demographic data, inpatient and outpatient hospital visits, inpatient rehabilitation visits, contract-medical care, outpatient drug prescriptions, therapeutic services, and care data. For the purpose of data profiling we used the open-source OHDSI tool WhiteRabbit[2] [7]. WhiteRabbit analyzes the provided source data and automatically generates a scan report including detailed information about tables, their data elements and data types as well as frequency

[2] https://github.com/OHDSI/WhiteRabbit

distributions of source values. The scan report can be used as a starting point for further comparison to German clinical data and feasibility analysis.

2.2. Comparison with clinical data mapping

As stated by Henke et al. [8] there is an intersection between German claims data and clinical data with respect to demographic, visit, procedure and diagnosis data from inpatient hospital care. Overlapping data elements are then focused on to assess mapping similarities between German claims data and clinical data. For the mapping of clinical data, we used the documentation of the Extract-Transform-Load (ETL) processes implemented by Peng et al. [5] and Zoch et al. [6] that focuses on mapping clinical data from German university hospitals to OMOP CDM [5] and mapping German claims data to OMOP CDM restricted to inpatient billing data [6], respectively.

2.3. Feasibility of harmonizing claims data in OMOP CDM

The final step of our concept took the results from the data profiling and the comparison with clinical data mapping into account to analyze the feasibility of harmonizing German claims data in OMOP CDM v5.3.1. First, we identified all vocabularies used in German claims data. Next, we evaluated the presence of required vocabularies in the OHDSI vocabulary web application ATHENA [9] to map the source data. Moreover, we examined the ETL processes for existing source code mappings to Standard OMOP Concepts through the source_to_concept_map table in OMOP CDM (interim mappings). To measure the current vocabulary coverage of German claims data in OMOP CDM, we categorized the source vocabularies into “available in ATHENA”, “available through interim mapping” and “not available”. After considering the vocabularies, we evaluated the feasibility of mapping source data elements to the standardized OMOP CDM fields. As part of this, the influence of any missing vocabularies on the mapping of data elements was included. For the measurement of the mapping coverage, we distinguished between the categories “possible”, “missing vocabularies”, “missing OMOP CDM fields” and “missing vocabularies and OMOP CDM fields”.

3. Results

3.1. Vocabulary

Based on the categories assigned to the source vocabularies, we calculated the vocabulary coverage in OMOP CDM in percentage (see supplementary file[3] for details). Figure 2 shows that only 15% of the source vocabularies are available in OMOP CDM but 55% are not. Missing vocabularies mainly concern vocabularies for therapeutic services and their indications, billing vocabularies such as the German Uniform Assessment Standard (EBM[4]) or services provided by psychiatric institutional outpatient

[3] https://caruscloud.uniklinikum-dresden.de/index.php/s/cs4HCB7LkJyjaFK

[4] Einheitlicher Bewertungsmaßstab

clinics (PIA[5]). Nevertheless, the remaining 30% of the source vocabularies can be mapped to standard OMOP CDM vocabularies through interim mappings.

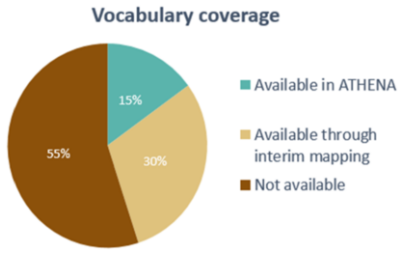


Figure 2. Vocabulary coverage of German claims data in OMOP CDM

3.2. Mapping

In order to gain knowledge about the current feasibility of mapping German claims data to OMOP CDM, we calculated a mapping coverage based on the flags mentioned in Section 2.3. As shown in Figure 3, 87% of the data elements can already be mapped to OMOP CDM. Mapping is currently not possible for 13% of the data elements. Reasons for this are missing vocabularies (8%), missing OMOP CDM fields (3%) or both, missing vocabularies and OMOP CDM fields (2%). In this context, we identified that the cost table in OMOP CDM is solely constructed for US specific hospital charges leaving no possibilities to map German charge data.

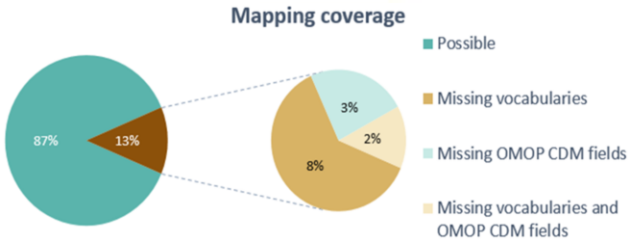


Figure 3. Mapping coverage of German claims data in OMOP CDM

4. Discussion and Conclusions

The results show a high mapping coverage of German claims data in OMOP CDM. However, we identified some obstacles regarding missing vocabularies and OMOP CDM fields, resulting in a mapping coverage lower than 90%. When reviewing the literature, our results are comparable to findings in other countries [10–14]. During our next steps, we address the current obstacles to prevent data loss during the transformation of German claims data to OMOP CDM. First, we focus on the preparation of missing vocabularies and their integration in the standardized vocabulary of OMOP CDM. Afterwards, we develop an approach to handle missing OMOP CDM fields and implement an ETL process to transform German claims data to OMOP CDM including

[5] Psychiatrische Institutsambulanzen

a qualitative control mechanism for the prepared vocabularies and the semantic mapping of source data elements to OMOP CDM. Consequently, we are building the basis for making German claims data available for international research as well as for the linkage with clinical data in OMOP CDM.

Declaration

Conflict of Interest: The authors declare that there is no conflict of interest.

Author contributions: All authors contributed substantial ideas and participated in editing and revising of the manuscript. All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

Funding: The study was funded by the German Innovations Fund of the Federal Joint Committee in Germany (G-BA) (grant number: 01VSF20013).

References

[1] Semler SC, Wissing F, Heyder R, German Medical Informatics Initiative. Methods Inf. Med. 2018;57:e50-6.
[2] Green LA, Fryer Jr GE, Yawn BP, Lanier D, Dovey SM. The ecology of medical care revisited. New England Journal of Medicine. 2001 Jun 28;344(26):2021-5. doi:10.1056/NEJM200106283442611.
[3] Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. Journal of biomedical informatics. 2016 Dec 1;64:333-41. doi:10.1016/j.jbi.2016.10.016.
[4] Data Standardization – OHDSI, (n.d.). https://www.ohdsi.org/data-standardization/ (accessed November 7, 2022).
[5] Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. International Journal of Medical Informatics. 2023 Jan 1;169:104925.
[6] M. Zoch, E. Henke, I. Reinecke, Y. Peng, R. Gebler, M. Gruhl, and M. Sedlmayr, Extract, Transform and Load German Claim Data to OMOP CDM – Design and Implications, in: German Medical Science GMS Publishing House, 2022: p. DocAbstr. 153. doi:10.3205/22gmds057.
[7] White Rabbit, (n.d.). http://ohdsi.github.io/WhiteRabbit/WhiteRabbit.html (accessed November 11, 2022).
[8] E. Henke, M. Zoch, F. Bathelt, M. Spoden, T. Ruhnke, C. Günster, and M. Sedlmayr, Identifikation von Gemeinsamkeiten klinischer Daten und GKV-Routinedaten für deren Verknüpfung in OMOP CDM zur stationären Qualitätsmessung, in: German Medical Science GMS Publishing House, 2022: p. DocAbstr. 85. doi:10.3205/22gmds007.
[9] Athena, (n.d.). https://athena.ohdsi.org/search-terms/start (accessed November 25, 2022).
[10] Haberson A, Rinner C, Schöberl A, Gall W. Feasibility of mapping austrian health claims data to the OMOP common data model. Journal of Medical Systems. 2019 Oct;43:1-5. doi:10.1007/s10916-019-1436-9.
[11] Kim H, Choi J, Jang I, Quach J, Ohno-Machado L. Feasibility of representing data from published nursing research using the OMOP common data model. InAMIA Annual Symposium Proceedings 2016 (Vol. 2016, p. 715). American Medical Informatics Association..
[12] Cho S, Sin M, Tsapepas D, Dale LA, Husain SA, Mohan S, Natarajan K. Content coverage evaluation of the OMOP vocabulary on the transplant domain focusing on concepts relevant for kidney transplant outcomes analysis. Applied Clinical Informatics. 2020 Aug;11(04):650-8. 650–658. doi:10.1055/s-0040-1716528.
[13] Lamer A, Depas N, Doutreligne M, Parrot A, Verloop D, Defebvre MM, Ficheur G, Chazard E, Beuscart JB. Transforming French electronic health records into the Observational Medical Outcome Partnership's common data model: a feasibility study. Applied Clinical Informatics. 2020 Jan;11(01):013-22.. doi:10.1055/s-0039-3402754.
[14] Sathappan SM, Jeon YS, Dang TK, Lim SC, Shao YM, Tai ES, Feng M. Transformation of electronic health records and questionnaire data to OMOP CDM: a feasibility study using SG_T2DM Dataset. Applied Clinical Informatics. 2021 Aug;12(04):757-67. doi:10.1055/s-0041-1732301.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230054

Can Patient Contributed Data (PCD) Leverage Connected Health Technology for Cardiac Rehabilitation in Austria?

Rada HUSSEIN[a,1], Adrienne PICHON[b], Jan OLDENBURG[c], Mahdi SAREBAN[a,d] and Josef NIEBAUER[a,d]
[a] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[b] *Department of Biomedical Informatics, Columbia University, New York, USA*
[c] *Participatory Health Consulting, LLC, Virginia, USA*
[d] *University Institute of Sports Medicine, Prevention and Rehabilitation, Paracelsus Medical University, Salzburg, Austria*
ORCiD ID: Rada Hussein https://orcid.org/0000-0003-1257-4848, Adrienne Pichon https://orcid.org/0000-0002-9393-3410, Jan Oldenburg https:https://orcid.org/0000-0001-5193-9382, Mahdi Sareban https://orcid.org/ https://orcid.org/0000-0002-8146-0505, Josef Niebauer https://orcid.org/0000-0002-2811-9041

Abstract. New technologies such as devices, apps, smartphones, and sensors not only enable people to self-monitor their health but also share their health data with healthcare professionals. Data collection and dissemination occur across a wide variety of environments and settings, tracking everything from biometric data to mood and behavior, which has been termed Patient Contributed Data (PCD). In this work, we created a patient journey, enabled by PCD, to shape a connected health model for Cardiac Rehabilitation (CR) in Austria. Consequently, we highlighted the potential PCD benefit, which is a postulated increasing uptake of CR and improved patient outcomes through apps in a home-based setting. Finally, we addressed the related challenges and policy barriers that hinder the implementation of CR-connected health in Austria and identified actions to be taken.

Keywords. Cardiac rehabilitation, connected health, interoperability, patient-contributed data, patient journey

1. Introduction

Patient Contributed Data (PCD) is defined as "*Any data, information, or insights created, collected by, or originating from a person regarding his or her health and care. It is particularly relevant when shared with one or more clinical care team members for the purpose of collaboration around the person's health*" [1]. PCD can play an essential role in linking connected health technologies, such as: digital health, eHealth, mHealth, telehealth, telemedicine, telecare, remote care, and assisted living. Connected health is understood as "*a socio-technical model for healthcare management and delivery by using technology to provide healthcare services remotely*" [2]. Thus, it includes all of

[1] Corresponding Author: Rada Hussein, Ludwig Boltzmann Institute for Digital Health and Prevention, Lindhofstrasse 22, 5020 Salzburg, Austria; E-mail: rada.hussein@dhp.lbg.ac.at

the systems, processes, personnel and technology used to meet a specific patient care goal. Recent studies highlight the benefits of connected health for cardiovascular disease prevention and management [3], via providing activity tracking, blood pressure monitoring, dietary interventions, smoking cessation, lipid management, and risk assessment. Consequently, connected health expands Cardiac Rehabilitation (CR) services by supporting home-based programs as an effective adjacent or alternate approach to center-based CR programs [4,5]. Notably, there are many barriers that hinder a wider uptake of CR [6], including lack of referral, digital literacy, and reimbursement, as well as lack of a legal framework for integration into the Austrian Electronic Health Record (ELGA), and further concerns with data protection.

2. Methods

To explore potential PCD benefits in providing connected health for CR in Austria [7], we used a patient journey method [8,9] to highlight the role of PCD in addressing and overcoming the current CR challenges.

2.1. Markus' Journey: Out-patient and home-based CR with rich data capture

During this patient journey (see Figure 1), Markus aims to better control his cardiovascular risk factors, i.e., mainly high blood pressure, cholesterol, and physical inactivity to reduce the risk of disease progression and future cardiovascular events. As a result, he is motivated to pursue further healthy behavior changes.

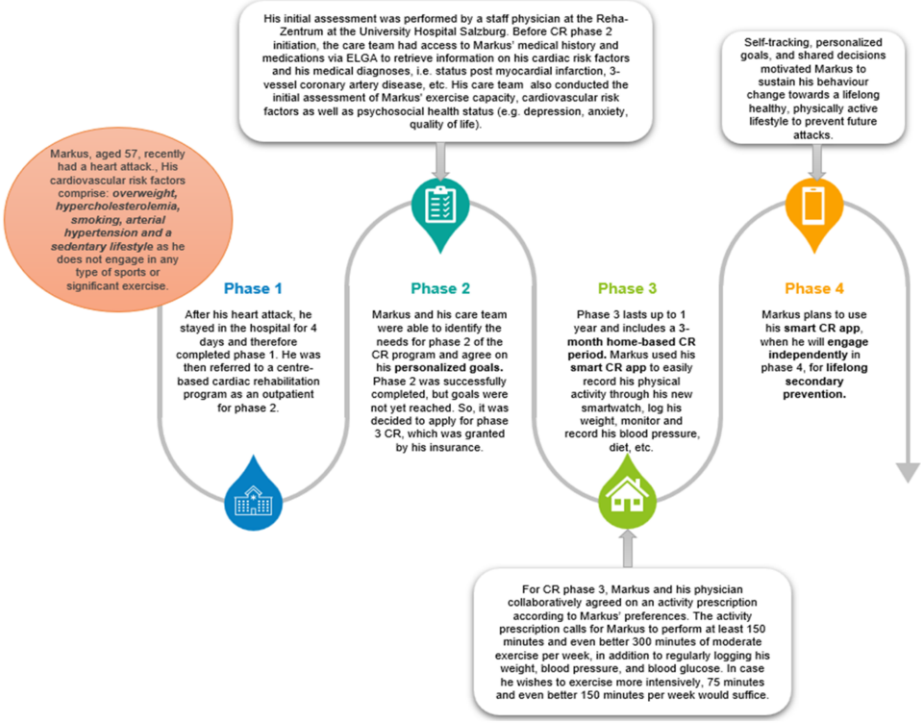


Figure 1. Markus' Journey during the four phases of cardiac rehabilitation.

Although Markus is not very engaged with technology, he finds a sense of satisfaction in using the smart CR app, recommended by his rehabilitation team, to track his progress and communicate with his care team during the four phases of CR in Austria [7].

2.2. Data Journey and used apps

The smart CR app enabled Markus to easily record his physical activity through his new smartwatch, log his weight, monitor and record his blood pressure (with the ability to link to a Bluetooth connected device and streamline blood pressure readings straight into the app). He also logged his daily caloric intake and dietary content of fat, saturated fat, sodium, and other nutrients in addition to eating habits. Tracking his diet was the hardest for Markus and he was pretty sporadic about doing so. His wife and daughter helped him in tracking these activities, entering the measurements when he complained about the burdensome tracking. Also, Markus worked on improving his skills in using the CR app, other health apps and his smartwatch.

On a weekly basis, all recorded data from Markus' smartwatch and smart CR app were transferred to his physician via passive data sharing. His weight from the app connected to his digital scale was automatically transferred (passive sharing). Markus had to select his mood indicator and enter other related symptoms via the smart CR app (active sharing). The physician requested all these data to assess Markus' risk factors during CR phase 2 (solicited) after working together to make a plan for tracking. Because of this collaboration, his clinician was able to assess Markus's risk factors weekly and make relevant shared-decisions on the required intervention plan and/or education. For example, when Markus did not achieve the physical activity goals one week, his rehab team was able to customize the activity plan for the following week to help him fulfill the World Health Organization (WHO) activity requirement, i.e., perform at least 150-300 min a week of moderate-intensity or 75-150 min a week of vigorous-intensity aerobic physical activity or an equivalent combination thereof. In addition, the multidisciplinary rehab team was able to share educational material on sustaining healthy lifestyles with Markus when he ran into barriers trying to lose weight during the three weeks. Markus also received educational material that helped him with smoking cessation. Figure 2 summarizes the used apps and aligned data journey. An extended discussion on PCD use in cardiac rehabilitation can be found in [1].

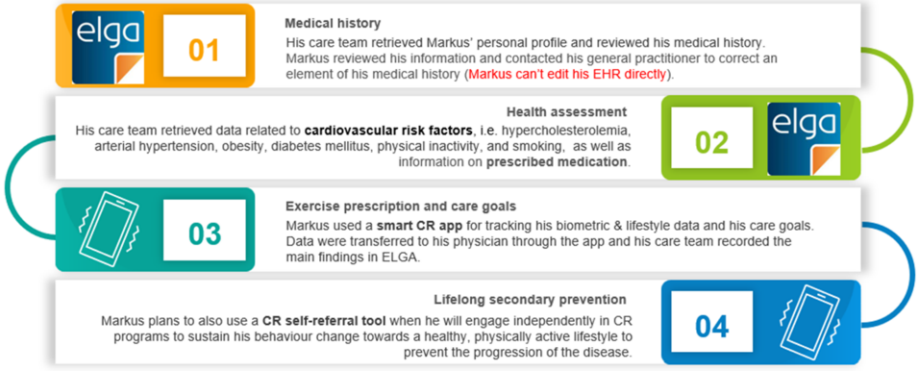


Figure 2. Markus' data journey and used apps.

3. Results and Discussions

The patient journey revealed the importance of PCD as a cornerstone in realizing a CR connected health model. Moreover, it identified current gaps, challenges in existing systems and relevant workflows, and the need for CR apps, as follows:

- The existing law of ELGA does not give the right to patients to write PCD into their records. Patients need to approach their care team who should be able to grant permission to upload the PCD on patients' behalf.
- Smart CR apps with proper educational material based on established behavioral change techniques are required to empower and engage patients, especially those with limited digital skills.
- Self-Referral apps can be introduced to overcome the challenge of the low referral and uptake rate of CR programs.

Subsequently, we addressed these challenges and recommended actions to be taken for realizing the CR connected health model in Austria as listed in Table.1.

Table 1. Challenges facing CR connected health model [6] and recommended actions to be taken

Policy Barriers	Ethical Concerns	Societal Factors	Technical Challenges
- Align the current national policies with the European Health Data Space (EHDS) regulations [10].	- Tackle the issues that hinder the uptake of digital health, e.g., data ownership, protection, sharing, and control. That's where the EHDS comes into play.	- Expand the existing programs of digital health literacy.	- Extend the infrastructure and interoperability layer for integrating (or linking) PCD with ELGA [11].
- Develop and finance a transparent certification process for CR apps.	- Support transparent, understandable, explainable, and fair use of Artificial Intelligence (AI).	- Provide easy access to digital platforms, e.g., single ID.	- Enable dynamic informed consent mechanisms.
- Harmonize the financing and cost for reimbursement policies.	- Provide equitable access to data within the connected health model.	- Promote knowledge about efficacy of home-based CR solutions.	- Develop AI automation tools and CR smart apps [12].

4. Conclusions

The described patient-centered connected healthcare model for CR integrates electronic health record, digital health technologies and cardiac telerehabilitation through PCD. This enables CR patients to stay in their home environment and capture rich data that supports them in reaching their health goals and has the potential to support their long-term priorities. These data are shared with the care team to monitor patients' cardiovascular risk factors and understand how patients are progressing. This also has a direct benefit to CR patients through improving the quality and outcomes of their care plan, which can be expected to translate into reduced morbidity and mortality. Indeed, it has been previously shown, that heart attack survivors who completed CR were 40% less likely to experience another heart attack [13]. In order to increase the number of patients who enroll in CR as well as to improve and individualize the quality of CR programs,

we need to overcome addressed challenges and policy barriers, which currently hinder the alignment with EHDS requirements [10].

5. Acknowledgments

This work was conducted as one of the patient journeys of the PCD whitepaper, sponsored by the Health Level Seven (HL7®) Patient Empowerment workgroup. The authors thank other members of the PCD group for their valuable comments and fruitful discussions.

References

[1] HL7 Patient Empowerment Workgroup. Patient Contributed Data (PCD) White Paper [Internet]. 2022. Available from: http://www.hl7.org/documentcenter/public/ballots/2022SEP/downloads/HL7_PCD_R1_I1_2022SEP.pdf
[2] Nordic Innovation. Branding Nordic Healthcare Strongholds. A Nordic Story About Smart Digital Health [Internet]. 2018. Available from: https://norden.diva-portal.org/smash/get/diva2:1297054/FULLTEXT01.pdf
[3] Wongvibulsin S, Martin SS, Steinhubl SR, Muse ED. Connected Health Technology for Cardiovascular Disease Prevention and Management. Curr Treat Options Cardiovasc Med. 2019 May 18;21(6):29.
[4] Epstein E, Patel N, Maysent K, Taub PR. Cardiac Rehab in the COVID Era and Beyond: mHealth and Other Novel Opportunities. Curr Cardiol Rep. 2021 Mar 11;23(5):42.
[5] Antoniou V, Davos CH, Kapreli E, Batalik L, Panagiotakos DB, Pepera G. Effectiveness of Home-Based Cardiac Rehabilitation, Using Wearable Sensors, as a Multicomponent, Cutting-Edge Intervention: A Systematic Review and Meta-Analysis. J Clin Med. 2022 Jun 29;11(13):3772.
[6] Scherrenberg M, Wilhelm M, Hansen D, Völler H, Cornelissen V, Frederix I, et al. The future is now: a call for action for cardiac telerehabilitation in the COVID-19 pandemic from the secondary prevention and rehabilitation section of the European Association of Preventive Cardiology. Eur J Prev Cardiol. 2020 Jul 2;2047487320939671.
[7] Niebauer J. Cardiac rehabilitation in Austria. Wien Med Wochenschr. 2018;168(1):46–9.
[8] Simonse L, Albayrak A, Starre S. Patient journey method for integrated service design. Design for Health. 2019 Jan 2;3(1):82–97.
[9] Bolz-Johnson M, Meek J, Hoogerbrugge N. “Patient Journeys”: improving care by patient involvement. Eur J Hum Genet. 2020 Feb;28(2):141–3.
[10] Machalska M. EACH Plan for Cardiovascular Health Launched [Internet]. EACH. 2022 [cited 2022 Nov 23]. Available from: https://www.cardiovascular-alliance.eu/each-plan-for-cardiovascular-health-launched/
[11] Hussein R, Crutzen R, Gutenberg J, Kulnik ST, Sareban M, Niebauer J. Patient-Generated Health Data (PGHD) Interoperability: An Integrative Perspective. EFMI Medical Informatics Europe; 2021; Athens.
[12] Sotirakos S, Fouda B, Mohamed Razif NA, Cribben N, Mulhall C, O’Byrne A, et al. Harnessing artificial intelligence in cardiac rehabilitation, a systematic review. Future Cardiol. 2022 Feb;18(2):154–64.
[13] Cardiac Rehabilitation | CardioSmart – American College of Cardiology [Internet]. CardioSmart. [cited 2022 Nov 23]. Available from: http://www.cardiosmart.org/topics/cardiac-rehabilitation

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230055

Towards a Clinically Meaningful Model to Structure the Development of Interoperable Order Sets, Applicable to the Point of Care in Any EMR

Mobin YASINI[a,1] Dennis RAUSCH[a], Lise MARIN[a], Ingrid HOCHHEIM[a], Navinder SINGH DHILLON[a], and Michael DAHLWEID[a]
[a] *Dedalus Research Lab, Italy*
ORCiD ID: Mobin YASINI https://orcid.org/0000-0002-9281-8875

Abstract. Standardized order sets are a pragmatic type of clinical decision support that can improve adherence to clinical guidelines with a list of recommended orders related to a specific clinical context. We developed a structure facilitating the creation of order sets and making them interoperable, to increase their usability. Various orders contained in electronic medical records in different hospitals were identified and included in different categories of orderable items. Clear definitions were provided for each category. A mapping to FHIR resources was performed to relate these clinically meaningful categories to FHIR standards to assure interoperability. We used this structure to implement the relevant user interface in the Clinical Knowledge Platform. The use of standard medical terminologies and the integration of clinical information models like FHIR resources are key factors for creating reusable decision support systems. The content authors should be provided with a clinically meaningful system to use in a non-ambiguous context.

Keywords. Medical Record, Standing Orders, Electronic Prescribing

1. Introduction

One of the potentials of digitalizing medical records is the ability to implement clinical decision support systems (CDSS) that could help healthcare providers in a variety of decisions and patient care tasks [1]. CDSS are software applications designed to assist health care professionals in decision making throughout the care process. When used at the point of ordering, CDSS can integrate evidence-based clinical guidelines with computerized physician order entry systems (CPOE) [2].

Order sets are clinical tools that deliver guidance by providing a list of recommended orderable items applicable in a specific clinical context. A clinical context is a combination of various conditions including disease, symptoms, comorbidities, stage of the problem, stage of the care, demographic patient characteristics, etc.[3] that define the applicability of an order set. Orderable items could be of various types (medication, lab test, imaging, procedure, etc.) and each type could have various attributes (timing, count,

[1] Corresponding Author: Mobin Yasini, Dedalus France, 22 Avenue Galilée, 92350 Le Plessis-Robinson, France; E-mail: mobin.yasini@dedalus.eu

frequency, etc.). Standardized order sets can positively influence adherence to evidence-based guidelines and could be used to ensure appropriate orders are being made by health professionals and therefore reduce errors [4].

Despite ongoing development and promising potentials of CDSS in general, lack of interoperability renders many CDSS as cumbersome stand-alone systems that cannot communicate effectively with other systems [1] or may be used only in the EMRs for which they had been developed. This makes the applicability of CDSS including order sets limited to local use. Interoperability enables better workflows and reduces ambiguity among systems and could be used in various areas. For example, interoperable EMRs allow the electronic sharing of patient information between different systems and healthcare providers, improving the ease with which doctors can provide care to their patients. Interoperable CDSS could disseminate the knowledge that they represent in various systems and institutions and reduces the implementation costs through their reusability. The use of standardized formalisms for knowledge representation, like terminologies as well as the integration of semantically enriched clinical information models, contributes to the development of interoperable CDSS [5,6].

We have already presented the Clinical Knowledge Platform (CKP) as an ecosystem in which clinicians could create and share interoperable order sets [3]. In this paper, we present the method that we used to define various orderable items and the relevant attributes specific to each orderable item in a clinically meaningful manner. This provides clinicians with an easily understandable platform with necessary items that enables them to create interoperable order sets.

2. Methods

We need to provide the order set authors with a tool that empowers them to find the orderable items easily and rapidly, together with the relevant attributes. For that purpose, a working group including health professionals and experts in information technology (IT) and clinical informatics was created.

Defining the classification of orderable items: In order to identify the various clinically meaningful categories of orderable items, the health professionals of the working group have studied a set of 325 medical records selected randomly in different general hospitals in France and Germany. Medical records were first randomly assigned to two groups. The first group, the study group, containing two thirds of these records, was used to structure the categories of orderable items. The remaining third was used to validate the categories obtained. The orders of each medical record in the study group were listed. These orders were then analyzed to identify the categories to which they belong. Whenever an order could not be included in a category already created, a new category was added. The categories of orders found in the medical records were therefore enumerated incrementally, as they were discovered. A clear clinical definition was assigned to each category of orderable items to define the meaning and functionalities of each category.

Defining the attributes of each category: We then analyzed the orders of each category, to identify the various attributes that a category could include. For example, an order for the prescription of a drug may have attributes such as pharmaceutical form, route of administration, duration of administration, frequency, etc.

Mapping to FHIR: CKP uses FHIR resources and standard interrelated multi-terminology servers [3] to assure interoperability. FHIR resources for orderable items

were investigated by the working group to map the relevant resources to the founded categories. Then, the attributes for each category were discussed and linked to the relevant FHIR content.

Developing the order set designer in the CKP: The categories of orderable items and the relevant attributes were then used to develop the user interface in the CKP environment. Various catalogs of orderable items were tagged to the relevant categories and implemented in the CKP database. The content creator could search for orderable items and assemble the desired order set.

3. Results

13 clinically meaningful categories and 40 attributes were identified for orderable items. All the categories of orderable items found in the medical records, together with the mappings to FHIR resources and the definitions are represented in table 1.

Table 1. Categories of orderable items, definitions, and mapping to FHIR resources

Orderable category	Definition	FHIR correspondence
Imaging/nuclear medicine	Medical imaging refers to several different technologies that are used to view the human body in order to diagnose, monitor, or treat medical conditions. It includes radiography, CT, Fluoroscopy, Mammography, MRI, Ultrasound. Nuclear medicine uses radioactive material inside the body to see how organs or tissue are functioning (for diagnosis) or to target and destroy damaged or diseased organs or tissue (for treatment).	ServiceRequest
Medication/ vaccination/ Blood & Blood derivatives	A substance that is taken into or placed on the body for a specific medical purpose (cure a disease, treat a medical condition, relieve symptoms, prevent disease, etc.), Ordering vaccines and whole blood and blood components including RBC, Platelets, Plasma, Packed cells, etc. are also categorized in this group	Medication Request
Microbiology, lab tests, genetics & pathology	A medical procedure that involves testing a sample of blood, urine, or other substance or tissue from the body. These tests can help determine a diagnosis, plan treatment, check to see if treatment is working, or monitor the disease over time	ServiceRequest
Nutrition & Diet	Oral diets (including general diets such as general Healthy diet, or therapeutic diets such as consistent carbohydrate, 2-gram Sodium, or fluid restricted), oral nutrition supplements (such as nutritionally complete pre-packed drinks), enteral nutrition (tube feedings) and infant formula which govern the distribution of food and nutritional products used to feed patients within an in-patient setting.	NutritionOrder
Consultation	Asking the opinion of another specialist consultation and assessments about the patient	ServiceRequest Appointment
Nursing	Any activity that is motivated by the intention to provide Nursing care (care procedures, positioning, feeding, washing, shaving, talking to the patient etc.)	ServiceRequest
Lines & drains	Any activity that results in placing, removal or maintenance of Lines & Drains (urinary catheter, NG tube, airway, probes, etc.)	ServiceRequest
Tracking notification	A request to convey information, e.g., the physician or a CDS system proposes that an alert be sent to a responsible provider, the CDS system proposes that the public health agency be notified about a reportable condition.	Communication Request
Surgery	A medical act that is performed for the purpose of structurally altering the human body by incision or destruction of tissues as a part of the practice of medicine. It includes operations, but also activities like installing an ilizarov frame, pacemaker implants, etc.	ServiceRequest

Allied care	Allied care includes anything not done by physicians or nurses (physiotherapy, speech, behavioral, etc.)	ServiceRequest
Devices, gear, prostheses	Ordering a tool, device, prothesis to be applied for the patient including mattresses, braces, ilizarov frame, pacemaker, etc.	DeviceRequest SupplyRequest
Wound care	Specific procedures for wounds (bandage, debridement, etc.)	ServiceRequest
Procedure	Other types of acts used to diagnose, measure, or treat problems such as disease or injury. It includes radiotherapy, tracheotomy, ECG, EEG, pacemaker check, endoscopy, biopsies, LP, PFT, etc.	ServiceRequest

Each attribute could be related to one or more categories. For example, an attribute like the day of the week which, if provided, specifies that the action happens only on the specified day, is related to all categories of orderable items. However, the attribute route of administration is related only to medication and nutrition categories. We mapped each category to the relevant HL7 FHIR resource. Various categories were mapped to service request resource which is not necessarily understood by clinicians.

These categories were developed from two thirds of the studied medical records and were validated on the remaining third. The validation showed that the data model was able to represent all the orders contained in the remaining third of the medical records.

Figure 1 shows some superposed screenshots of the implementation in the CKP.

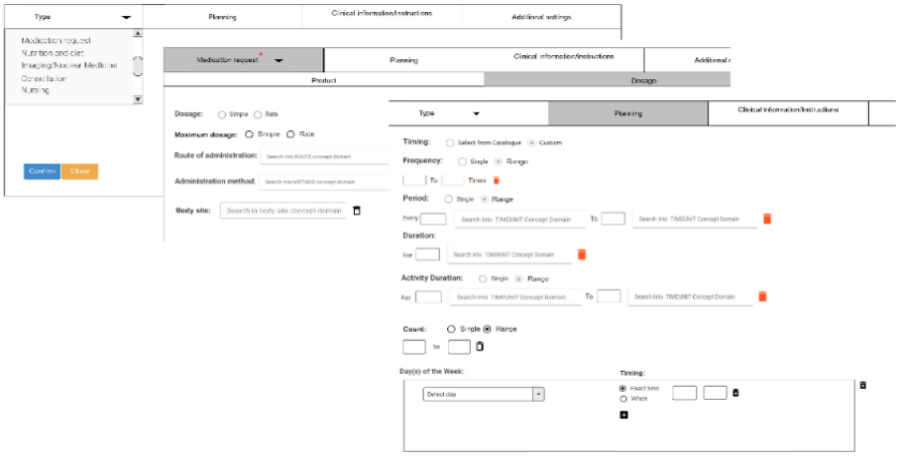


Figure 1. When the category is selected to medication request, the relevant attributes are automatically presented to the user. The attributes are also allocated to some groups to provide the user with a better usability.

While creating an order set, the user could directly search the orderable item or find it by choosing the relevant category of orderable items. Then, according to the category to which the orderable item belongs, the relevant attributes will be displayed to the user to specify the exact application of the orderable item in question.

4. Discussion

In this study, we have proposed 13 clinically meaningful categories of orderable items together with clear definitions and relevant attributes specific to each item. The content providers and clinicians could therefore create various order sets in the CKP. Using FHIR resources and mapping clinical concepts to standard terminologies in the CKP makes these order sets interoperable and allow them to be shared with various systems.

Our approach to the use of standardized formalisms for knowledge representation as terminologies as well as the integration of semantically enriched clinical information

models like FHIR resources, corroborates with the results of other studies that consider these issues as key factors for reusable CDSS [5]. Standardized order sets are shown to be associated with reduced hospital stays, decreased adverse patient effects, lower risk of mortality and increased cost-effectiveness [6–8].

Only health professionals were involved in the phase of defining the categories because the output had to be clinically meaningful. General hospitals were selected because all specialties had to be covered. However, in most hospitals, the orders are already organized by categories that may have a bias on the categories of orderable items that are used while developing an order set. The working group has merged some of the categories (like imaging and nuclear medicine). This was to improve the user experience while searching for orderable items. FHIR resources are complex, not clinically meaningful and include many attributes which are not necessarily understandable by clinicians. That's why we have mapped our categories to FHIR and gathered only functional attributes that are useful in our use case, instead of using original FHIR terms.

Further optimization of this structure, by other users than the developer team, including content providers, is required to confirm its generic nature. A quantitative evaluation of the impact of the interface in terms of the ease and intuitiveness to which physicians and content providers could create order sets would be of considerable interest. If the results of the evaluations are promising, recommended order items (within the categories) would be proposed as standards to be used for recommendations. CKP including the standardized and interoperable order sets can pave the way to other possible applications for improving the quality of medical practice. These applications may be based on various clinical contexts in which a physician prescribes or makes orders that are needed to be in accordance with the recommendations. The orderable categories can be also used to standardize clinical pathways, which involve a series of orders over time. Clinical pathways play a crucial role in patient care and standardizing them, like order sets, could lead to enhanced efficiency and better quality of care.

References

[1] Sutton RT, Pincock D, Baumgart DC, Sadowski DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ digital medicine. 2020 Feb 6;3(1):17.

[2] Forseen SE, Corey AS. Clinical decision support and acute low back pain: evidence-based order sets. Journal of the American College of Radiology. 2012 Oct 1;9(10):704-12.

[3] Dahlweid M, Rausch D, Hinske C, Darmoni S, Grosjean J, Santi J, Marin L, Yasini M. Clinical Knowledge Platform (CKP): A Collaborative Ecosystem to Share Interoperable Clinical Forms, Viewers, and Order Sets with Various EMRs. InDigital Professionalism in Health and Care: Developing the Workforce, Building the Future 2022 Aug 1 (pp. 117-121). IOS Press.

[4] Ansari S, Fung K, MacNeil SD, Nichols AC, Yoo J, Sowerby LJ. The use of standardized order sets to improve adherence to evidence-based postoperative management in major head and neck surgery. European annals of otorhinolaryngology, head and neck diseases. 2018 Sep 1;135(5):S107-11.

[5] Wulff A, Haarbrandt B, et al. An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR. Artificial Intelligence in Medicine. 2018;89:10-23.

[6] Kijsirichareanchai K, Ngamruengphong S, Rakvit A, Nugent K, Parupudi S. The utilization of standardized order sets using AASLD guidelines for patients with suspected cirrhosis and acute gastrointestinal bleeding. Quality Management in Healthcare. 2013 Apr 1;22(2):146-51.

[7] Fleming NS, Ogola G, Ballard DJ. Implementing a standardized order set for community-acquired pneumonia: impact on mortality and cost. The Joint Commission Journal on Quality and Patient Safety. 2009 Aug 1;35(8):414-AP5.

[8] Zeidan AM, Streiff MB, Lau BD, Ahmed SR, Kraus PS, Hobson DB, Carolan H, Lambrianidi C, Horn PB, Shermock KM, Tinoco G. Impact of a venous thromboembolism prophylaxis "smart order set": improved compliance, fewer events. American journal of hematology. 2013 Jul;88(7):545-9..

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230056

Application of Process Mining for Modelling Small Cell Lung Cancer Prognosis

Luca MARZANO[a,1], Sebastiaan MEIJER[a], Asaf DAN[b], Salomon TENDLER[b], Luigi DE PETRIS[b], Rolf LEWENSOHN[b], Jayanth RAGHOTHAMA[a] and Adam S. DARWICH[a]

[a] *Division of Health Informatics and Logistics, School of Engineering Sciences in Chemistry, Biotechnology and Health (CBH), KTH Royal Institute of Technology, Huddinge, Sweden.*

[b] *Department of Oncology-Pathology, Karolinska Institutet and the Thoracic Oncology Center, Karolinska University Hospital, Stockholm, Sweden*

Abstract. Process mining is a relatively new method that connects data science and process modelling. In the past years a series of applications with health care production data have been presented in process discovery, conformance check and system enhancement. In this paper we apply process mining on clinical oncological data with the purpose of studying survival outcomes and chemotherapy treatment decision in a real-world cohort of small cell lung cancer patients treated at Karolinska University Hospital (Stockholm, Sweden). The results highlighted the potential role of process mining in oncology to study prognosis and survival outcomes with longitudinal models directly extracted from clinical data derived from healthcare.

Keywords. Process mining, Real-world Data, oncology, small cell lung cancer, treatment decision

1. Introduction

Real-world data (RWD) from healthcare has the potential to inform real-world evidence of treatment effects. This includes time-to-event survival analysis to predict outcomes of administered therapies, and prognosis [1]. However, several aspects need to be considered when using RWD to inform decision-making, to avoid biases and extract robust results [2]. An important aspect is to consider the different treatment processes that the dataset represents [3]. In recent years, there has been an upswing in process mining applications within the healthcare domain [4]. Process mining is an approach that bridges process modelling and computational sciences, with the goal of extracting and analysing processes from the data. Process mining is usually carried out in several steps, encompassing process discovery, conformance check, and system enhancement [5]. This approach aims to improve the interpretation of outcomes generated from the data records

[1] Corresponding Author: Luca Marzano, Division of Health Informatics and Logistics, School of Engineering Sciences in Chemistry, Biotechnology and Health (CBH), KTH Royal Institute of Technology, Huddinge, Sweden, address: Hälsovägen 11, Huddinge, E-mail: lmarzano@kth.se .

produced during real-world processes [4]. Process mining in healthcare has mainly focused on production data, such as hospital clinical pathways, resource allocation, and scheduling [6]. Very little is currently known about process mining applied with the purpose of studying disease progression and survival outcomes. Oncology as a whole field, and small cell lung cancer (SCLC) in particular, is one of the diseases that would benefit most from this. Prognosis of oncological patients is mainly assessed with time-to-event analysis, such as Cox regression [7], with little consideration of treatment decision points over the course of the disease.

This paper showcases a process mining analysis of oncological survival data. The approach describes the design of a pipeline aimed to directly extract the processes from the clinical real-world database with the purpose of modelling longitudinally the prognosis and survival of the SCLC patients.

2. Methods

The data consisted of consecutive SCLC cases diagnosed and treated at the Karolinska University Hospital between 2008 and 2016 (n=705). The study was approved by the institutional review boards at Karolinska Institutet and at Stockholm County Council (2016/8-31). The present cohort was previously used to validate the eighth version of TNM (Tumour, Nodes, Metastases) staging system [8], study the prognostic impact of baseline patient characteristics [9], and the identification of subgroups of patients with similar survival times using unsupervised machine learning [3].

The type of treatment received by the patients in the cohort was chemotherapy (CT), chemotherapy and concomitant radiotherapy (CT+RT), radiotherapy (RT), and surgery (SR). No one received immunotherapy. Treatment choice is usually based on clinical variables, including patient's general condition (ECOG performance status), tumor extension (cancer stage) and response to the previous therapy as well as residual toxicity. According to the documented clinical or radiological progression of the tumor, the clinician may choose to change or reuse the same therapy. A rechallenge is considered as a new treatment line [9]. The progress free survival (PFS) for each line of therapy is defined as the interval between the start of the therapy and the earliest date of documented clinical or radiological progression according to standard clinical practice, or death.

Figure 1 shows the process mining workflow applied in this study. The main outcome of the pipeline is the longitudinal survival model extracted from the patients' records. Prior to extracting the processes, patient survival data were converted into an event-log format retrieving the PFS timestamps and the treatment decision follow up. After formatting, process mining was applied to extract the process map of the chemotherapy cycles with directly-follows graph technique [10]. This map could be defined as a directed graph, where nodes represent the selected therapy, and edges the follow up decision. Self-loops represented the re-challenge with the same therapy. Once the graph was extracted, the process behind PFS and treatment decision was studied. Following the querying of graphs related to patient groups of interest (e.g., patients with a specific tumor stage), process visualisation with metrics of interest were produced (edge frequency and node median PFS), sequence of the therapies, and transition matrices were provided, thus obtaining the longitudinal survival model developed directly from the patients' records.

All the analysis were carried out in R using the dplyr package for data processing, and bupaR package for process mining.

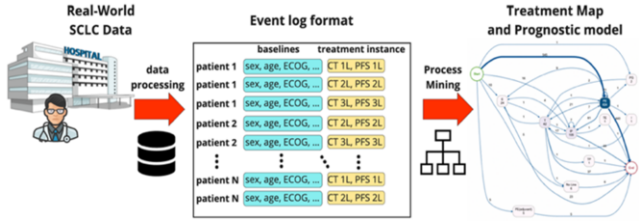


Figure 1. SCLC process mining workflow.

3. Results

A treatment decision event log with 1,622 instances was created from the patient data. The treatment map was extracted, and a dynamic process visualisation of the patients was produced detailing patient TNM staging. The developed pipeline in Figure 1 allowed to explore several cohorts by querying the subgraphs from the treatment map filtering the patients' characteristics of interest (such as cancer stage, and ECOG performance status). As example, Figure 2 shows the results for one of the cohorts: patient with IVB TNM stage (n=311). The 8th TNM classification introduced subclasses of stage IV SCLC patients, with most patients having multiple distant metastasis. The majority of SCLC cases are diagnosed with stage IVB and therefore the study of this case was of high clinical interest. From the longitudinal model we extracted the treatment-PFS graph, chemotherapy sequences, and treatment transition frequencies. The pipeline was tested also to extract longitudinal subgroup models for other TNM stages, for different treatment decisions (e.g., CT and CT+RT), and detected clusters in a previous study [3], thus showing the flexibility of the approach to different purposes.

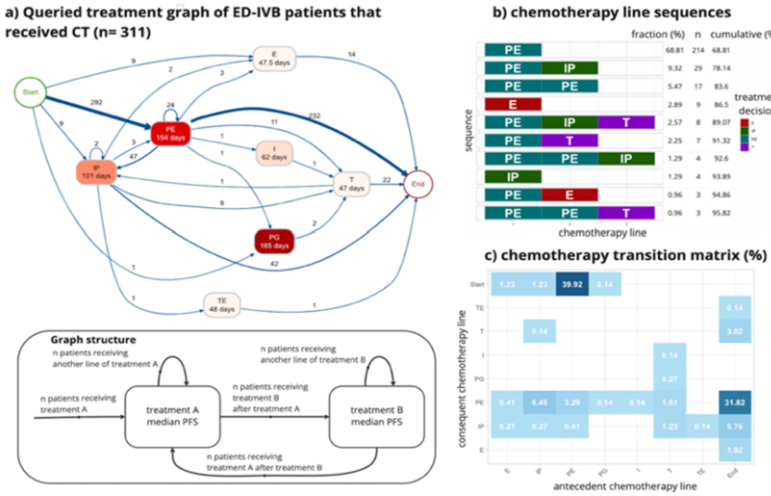


Figure 2. treatment process map, and chemotherapy (CT) line traces of ED-IVB patients. PE: platinum with etoposide, IP: platinum with irinotecan, T: topotecan, E: etoposide, I: irinotecan, TE: topetecan and etoposide, PG: platinum and gemcitabine

4. Discussion

The study presents an analysis of treatment pathways and associated outcomes in subgroups of SCLC patients treated with chemotherapeutic agents.

The chosen methodology was a pipeline that directly extracted treatment processes from the patient data, thus defining graphs from treatment and prognostic outcomes.

Results showed that process mining has potential for enriching the analysis of oncological cohorts by developing an increased understanding of the underlying treatment pathways and decision points. In addition to the rapid processing and the impactful visualisation, the technique can be used to inform longitudinal modelling of disease progression and subsequent impact on treatment decisions, as well as patient outcomes. The ready-to-use longitudinal models directly extracted from RWD constitute suitable objects for the design of multi-state survival models, or other process model solutions, such as causal networks [11].

To the best knowledge of the authors, this work constitutes the first application of process mining in healthcare that focuses on studying longitudinal survival outcomes in SCLC. In addition to process discovery, conformance check, and system enhancement, process mining could find reliable applications also for prognostic assessment and treatment evaluation.

This study has some limitations. For instance, the granularity of the data did not allow sufficient confidence to make inferences about the treatment effects. This is mainly due to the lack of longitudinal information in the retrospective cohort. On one hand, highly detailed information regarding the baseline patient and disease characteristics allowed the identification of patient groups of interest. On the other hand, follow-up information was limited to overall survival, PFS, and chemotherapy regimen. Further, most of the patients had extensive disease, TNM stage IVB, and short follow-up due to the early relapse during the first line therapy. Additional information on dose levels, adverse effects, and cancer progression would further improve the analysis. Increasing the sample size of the cohort, and the extension of the study involving multiple centres would benefit to the achievement of reliable evidence of SCLC studies.

Future work includes the collection of longitudinal information. Alternatives to the graph definition will be explored (e.g., adverse effects nodes, definition of edge weights, or new surrogate endpoints to study with the PFS). For what concerns the process mining, a larger variety of algorithms [10] will be tested to assess the variation and validity of the models. Involvement of the clinical experts in the model development will be a key factor to assess the clinical reliability of the approach, especially to leverage the gap between real processes and quality of the real-world data. Further considerations regarding the implementation challenges will be explored and discussed. Adherence with the current community standards for clinical pathways analytics (e.g., Observational Health and Data Science and Informatics [12]) and data interoperability in a large-scale infrastructure, such as the European Health Data Space [12], would be a key characteristic to implement.

5. Conclusions

Process mining applied to real-world healthcare data has the potential to allow visualisation of follow-up decisions in function of disease progression. The method can also be used to inform the design of multi-state models extracted directly from the data,

enabling more sophisticated statistical modelling of longitudinal data. The study provides insights into the role of process mining in oncology to study prognosis and survival outcomes, and additional indications on how to develop this further in the future.

References

[1] Schurman B. The Framework for FDA's Real-World Evidence Program. Appl Clin Trials 2019;28:15–7.

[2] Miksad RA, Abernethy AP. Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. Clin Pharmacol Ther 2018;103:202–5. https://doi.org/10.1002/CPT.946.

[3] Marzano L, Darwich AS, Tendler S, Dan A, Lewensohn R, de Petris L, et al. A novel analytical framework for risk stratification of real-world data using machine learning: A small cell lung cancer study. Clin Transl Sci 2022;15:2437–47. https://doi.org/10.1111/cts.13371.

[4] Munoz-Gama J, Martin N, Fernandez-Llatas C, Johnson OA, Sepúlveda M, Helm E, et al. Process mining for healthcare: Characteristics and challenges. J Biomed Inform 2022;127:103994. https://doi.org/10.1016/j.jbi.2022.103994.

[5] Martin N, de Weerdt J, Fernández-Llatas C, Gal A, Gatta R, Ibáñez G, et al. Recommendations for enhancing the usability and understandability of process mining in healthcare. Artif Intell Med 2020;109:101962. https://doi.org/10.1016/j.artmed.2020.101962.

[6] Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review. J Biomed Inform 2016;61:224–36. https://doi.org/10.1016/J.JBI.2016.04.007.

[7] de Neve J, Gerds TA. On the interpretation of the hazard ratio in Cox regression. Biometrical Journal 2020;62:742–50. https://doi.org/10.1002/bimj.201800255.

[8] Tendler S, Grozman V, Lewensohn R, Tsakonas G, Viktorsson K, De Petris L. Validation of the 8th TNM classification for small-cell lung cancer in a retrospective material from Sweden. Lung Cancer 2018;120:75–81. https://doi.org/10.1016/j.lungcan.2018.03.026.

[9] Tendler S, Zhan Y, Pettersson A, Lewensohn R, Viktorsson K, Fang F, et al. Treatment patterns and survival outcomes for small-cell lung cancer patients–a Swedish single center cohort study. Acta Oncol (Madr) 2020;59:388–94. https://doi.org/10.1080/0284186X.2019.1711165.

[10] van der Aalst W. Advanced Process Discovery Techniques. Process Mining 2016:195–240. https://doi.org/10.1007/978-3-662-49851-4_7.

[11] Krishnan SM, Friberg LE, Bruno R, Beyer U, Jin JY, Karlsson MO. Multistate model for pharmacometric analyses of overall survival in HER2-negative breast cancer patients treated with docetaxel. CPT Pharmacometrics Syst Pharmacol 2021;10:1255–66. https://doi.org/10.1002/PSP4.12693.

[12] European Commission D-G for H and FS. European Health Data Space 2022. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en#more-information (accessed February 1, 2023).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230057

Participatory Study to Explore Healthcare Professionals' Perceptions of a Connected Digital Solution for Adherence Monitoring of Recombinant Human Growth Hormone Treatment: Study Protocol and First Findings

Octavio RIVERA-ROMERO[a,1], Maria Felicia FAIENZA[b], Edoardo VERGANI[c], Chantal ZUCCHIATTI[d] and Ekaterina KOLEDOVA[e]
[a]*Electronic Technology Department, Universidad de Sevilla, Seville, Spain*
[b]*Department of Precision and Regenerative Medicine and Ionian Area, University of Bari Aldo Moro, Bari, Italy*
[c]*Department of Translational Medicine and Surgery, Agostino Gemelli Polyclinic Foundation, IRCCS – Catholic University of the Sacred Heart, Rome, Italy*
[d]*Training Specialist, Global Digital Health, Merck Healthcare KGaA, Darmstadt, Germany*
[e]*Global Medical Affairs Cardiometabolic & Endocrinology, Merck Healthcare KGaA, Darmstadt, Germany*
ORCiD ID: Octavio Rivera-Romero https://orcid.org/0000-0001-7212-9805

Abstract. Adherence to recombinant human growth hormone (r-hGH; somatropin, [Saizen®], Merck Healthcare KGaA, Darmstadt, Germany) treatment is fundamental to achieve positive growth outcomes in children with growth disorders and to improve quality of life and cardiometabolic risk in adult patients affected by GH deficiency. Pen injector devices are commonly used to deliver r-hGH but, to the authors' knowledge, none is currently digitally connected. Since digital health solutions are rapidly becoming valuable tools to support patients to adhere to treatment, the combination of a pen injector connected to a digital ecosystem to monitor treatment adherence is an important advance. Here, we present the methodology and first results of a participatory workshop that assessed clinicians' perceptions on such a digital solution – the aluetta™ smartdot™ (Merck Healthcare KGaA, Darmstadt, Germany) – combining the aluetta™ pen injector and a connected device, components of a comprehensive digital health ecosystem to support pediatric patients receiving r-hGH treatment. The aim being to highlight the importance of collecting clinically meaningful and accurate real-world adherence data to support data-driven healthcare.

Keywords. Adherence, connected device, digital ecosystem, recombinant human growth hormone, participatory research, pen injector

[1] Corresponding Author: Octavio Rivera-Romero, Electronic Technology Department, Universidad de Sevilla, Avda. Reina Mercedes s/n, 41012 - Seville, Spain; E-mail: orivera@us.es

1. Introduction

Adherence to medication is an important factor in the management of long-term conditions, including growth disorders such as growth hormone (GH) deficiency. Clinicians need to have an objective overview of their patient's adherence to treatment, allowing them to personalize treatment to achieve positive clinical outcomes [1]. Technology-based medication adherence monitoring is widely used [2], enabling the provision of clinically meaningful patient support based on the analysis of objective data. GH treatment is an example of a pharmacologic therapy for which adherence to daily injections is a key factor in achieving positive clinical outcomes [3].

The use of smartphones with Bluetooth-enabled adherence monitoring devices complementing medication injectors present an important opportunity for scalable adherence monitoring.[4, 5] Simple pen injectors are the most commonly used devices to deliver GH therapy; however, to the authors' knowledge, none is digitally-connected. The combination of the convenience of a pen injector and the benefits of a connected digital ecosystem will allow healthcare professionals (HCPs) to support patients to optimize treatment and clinical outcomes. One such example is the aluetta™ smartdot™ (Merck Healthcare KGaA, Darmstadt, Germany; Figure 1), combining the aluetta™ pen injector and a connected device. These are part of a large and comprehensive digital health ecosystem to support treatment with recombinant human GH (r-hGH; somatropin [Saizen®], Merck Healthcare KGaA, Darmstadt, Germany) and facilitate collaboration between patients/caregivers and HCPs to generate real-world data to support clinical decision making.[6]

Figure 1. aluetta™ smartdot™ connected device.

A typical approach to evaluate the usability of connected devices such as aluetta™ smartdot™ is to perform usability testing with a strong focus on maximizing safety. However, this approach has important limitations since it does not address other human factors such as intention to use, integration into a more comprehensive digital ecosystem, or trust in data capture.

As such, participatory health informatics research methods [7] are increasingly used to capture insights into the barriers to and opportunities for the integration of digital health solutions into clinical practice.[8, 9] Here, we present the methodology used in a workshop that explored clinicians' opinions and expectations concerning the aluetta™ smartdot™ to monitor adherence to r-hGH treatment in pediatric patients with growth disorders and adults affected by GH deficiency.

2. Methods

A pilot study following the workshop protocol was conducted in November 2022. Eight HCPs (five pediatric endocrinologists and three endocrinologists) from hospitals and universities from across Italy participated as members of the expert panel.

2.1. Workshop structure, activities, and materials

The workshop was co-moderated by two researchers with experience in participatory methods. Following an approach used in focus groups, participants were split into two teams with each team performing the activities in separate rooms. During the workshop (which was conducted in Italian and lasted 4 hours), participants were asked to provide their opinions of the aluetta™ smartdot™ prompted by a variety of predefined topics, based on their clinical experience.

The workshop consisted of five phases. *Phase 1* briefly outlined the project, introduced the moderators, described the Saizen® digital health ecosystem as an example, and outlined the general structure of the workshop. The objective of *Phase 2* was to understand the current healthcare context in Italy. Five topics were discussed independently by each team: the perceived importance of treatment adherence and potential factors impacting on it; the perceived usefulness of collecting adherence data; the current methods used to collect adherence data; HCPs' attitudes toward the use of digital health solutions for adherence monitoring; and the use of digital health solutions to support GH deficiency self-management. Participants identified relevant factors regarding each topic, and considered HCPs, patients/caregivers, and healthcare institutions/healthcare systems. Predefined templates were used to facilitate the activity (Figure 2).

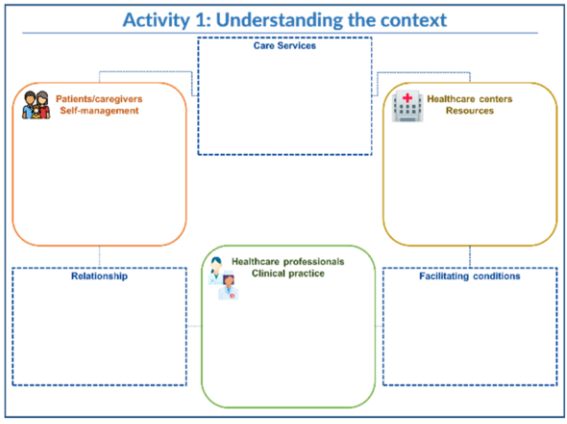


Figure 2. An example of the predefined templates.

The Saizen® digital health ecosystem was preselected and presented as a case study using an introductory video at the beginning of *Phase 3*. A prototype of aluetta™ smartdot™, which had the same look and feel, dimensions, weight, and ability to inject and store injection data as the final commercial product, but not enabled with complete functionality (e.g., the colors of the LED), was provided to each team. The aluetta™ pen injector was also provided. Participants assessed the physical characteristics of aluetta™ smartdot™, attempted to attach it to and detach it from aluetta™, assessed the perceived robustness of the attachment, and attempted to simulate an injection. Participants then

discussed several predefined topics focused on ergonomics, ease of use, perceived usefulness, and potential barriers and facilitators, in the context of GH deficiency management in Italy. Thereafter, the three most relevant topics from the given set were selected from the HCP's perspective, having identified relevant factors from their clinical experience.

Phase 4 aimed to explore HCPs' perceptions on how the technologic evolution of the device could impact on the adherence data collection process, HCPs' daily clinical practice, and patients' self-management. Several topics were predefined for each of these themes, and three scenarios were predefined: 1) Non-digital alternative (patients using a pen device to administer treatment and a paper diary to record their adherence; 2) Partially digital alternative (patients using a pen device to administer treatment and a digital diary integrated into a mobile app/website to record their adherence; and 3) Fully digital alternative (patients using aluetta™ smartdot™ and the growzen™ ecosystem that automatically registers their adherence data and transmits it to a secure internet-based cloud, as part of an adherence decision support system. Thus, enabling HCPs to monitor individual patients' adherence based on information retrieved from their aluetta™ pen injector).

In *Phase 5*, the relevant opinions raised during the previous activities were summarized and discussed by the participants. Members of each team selected the most relevant barriers to (and facilitators for) the use of aluetta™ smartdot™ in the current healthcare context in Italy. Thereafter, the participants discussed the level of agreement on their perceptions of aluetta™ smartdot™ according to the predefined topics.

2.2. Data collection and analysis

Data were collected via audio-recordings of the sessions, completion of the predefined templates, and facilitators' notes. Audio recordings were transcribed and translated into English for thematic analysis. Data from the templates and notes were combined and used to support the thematic analysis.

3. Discussion

The method employed to conduct the participatory workshop enabled collection of clinically valuable and understandable technology acceptancy information with regard to potential barriers to and facilitators for the use of aluetta™ smartdot™ to monitor treatment adherence in patients receiving r-hGH treatment in Italy. The workshop facilitated discussions about the perceived advantages of using aluetta™ smartdot™ to support data-driven health care, including the importance of unbiased data to support clinical decision making and personalized care. Our work reinforces emerging efforts to ensure comprehensive user involvement in the design and implementation of data-driven healthcare solutions.[10] While some topics required no explanation, others required clarification due to technical terminology (e.g., technologic and psychologic terms). This shows the importance of having an interdisciplinary support team during the participatory workshops. The predefined templates were understood and appropriately used by participants, allowing them to engage in the activities; this is aligned with previous experience on template-driven brainstorming (e.g., model business template).[11] Preliminary findings from the thematic analysis report that participants perceived aluetta™ smartdot™ to be highly useful and easy to use; the former driven by

the importance of automated collection of adherence data to improve data quality (more accurate, less risk of bias) and supporting clinical decision making.

4. Conclusions

We captured clinically meaningful insights on the usefulness of a new connected device – aluetta™ smartdot™ – to support adherence to r-hGH treatment. It is hoped that the outputs of this workshop will highlight the importance of collecting clinically meaningful, comprehensive, and accurate adherence data via a connected device to support clinical use and promote patient empowerment.

Acknowledgments and Conflicts of interest

The authors would like to thank all workshop participants, Luis Fernandez-Luque for his scientific support, and Amy Evans of inScience Communications, Springer Healthcare Ltd, UK for providing editorial assistance funded by Merck Healthcare KGaA, Darmstadt, Germany. OR-R has participated in advisory board for Merck. CZ and EK are employees of Merck Healthcare KGaA, Darmstadt, Germany, and EK also holds shares in the company.

References

[1] World Health Organization. Adherence to long-term therapies: evidence for action 2003. Available from: https://www.who.int/chp/knowledge/publications/adherence_report/en/.
[2] Aldeer M, Javanmard M, Martin RP. A review of medication adherence monitoring technologies. Appl Syst Innov. 2018;1(2):14.
[3] Fisher BG, Acerini CL. Understanding the growth hormone therapy adherence paradigm: a systematic review. Horm Res Paediatr. 2013;79(4):189-96.
[4] Hollenbach J, Simoneau T, Sun Y, Becene I, Almeida S, Langton C, et al. Design, methods, and baseline characteristics of a pilot, randomized, controlled trial of the effects of an electronic monitoring device on medication adherence in children with asthma. Contemp Clin Trials Commun. 2021 Mar;21:100706.
[5] Sun C, Malcolm JC, Wong B, Shorr R, Doyle MA. Improving glycemic control in adults and children with type 1 diabetes with the use of smartphone-based mobile applications: a systematic review. Can J Diabetes. 2019 Feb;43(1):51-8 e3.
[6] Savage MO, Fernandez-Luque L, Graham S, van Dommelen P, Araujo M, de Arriba A, et al. Adherence to r-hGH therapy in pediatric growth hormone deficiency: current perspectives on how patient-generated data will transform r-hGH treatment towards integrated care. Patient Prefer Adherence. 2022;16:1663-71.
[7] Denecke K, Gabarron E, Petersen C, Merolli M. Defining participatory health informatics - a scoping review. Inform Health Soc Care. 2021 Sep 2;46(3):234-43.
[8] Hu X, Fang H, Wang P. Factors affecting doctor's recommendation for mobile health services. Digit Health. 2022 Jan-Dec;8:20552076221125976.
[9] Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth. 2021 Mar 4;9(3):e21795.
[10] Zytko D, Wisniewski PJ, Guha S, Baumer EPS, Lee MK. Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains. Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, April 2022; Article No. 154; pp 1–4.
[11] Boßelmann S, Margaria T. Guided Business Modeling and Analysis for Business Professionals. In: Pfannstiel MA, Rasche C, editors. Service Business Model Innovation in Healthcare and Hospital Management: Models, Strategies, Tools. Cham: Springer International Publishing; 2017. p. 195-211.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230058

Utility-Preserving Anonymization in a Real-World Scenario: Evidence from the German Chronic Kidney Disease (GCKD) Study

Lisa PILGRAM[a,b,1], Elke SCHÄFFNER[c], Kai-Uwe ECKARDT[a,d], Fabian PRASSER[b], and GCKD Investigators
[a] *Department of Nephrology and Medical Intensive Care, Charité - Universitätsmedizin Berlin, Berlin, Germany*
[b] *Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Biomedical Innovation Academy, BIH Charité Junior Digital Clinician Scientist Program, Berlin, Germany*
[c] *Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany*
[d] *Department of Nephrology and Hypertension, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany*
ORCiD ID: Lisa Pilgram https://orcid.org/0000-0002-1020-0650
Elke Schäffner https://orcid.org/0000-0002-7925-4577,
Kai-Uwe Eckardt https://orcid.org/0000-0003-3823-0920
Fabian Prasser https://orcid.org/0000-0003-3172-3095

Abstract. Data sharing provides benefits in terms of transparency and innovation. Privacy concerns in this context can be addressed by anonymization techniques. In our study, we evaluated anonymization approaches which transform structured data in a real-world scenario of a chronic kidney disease cohort study and checked for replicability of research results via 95% CI overlap in two differently anonymized datasets with different protection degrees. Calculated 95% CI overlapped in both applied anonymization approaches and visual comparison presented similar results. Thus, in our use case scenario, research results were not relevantly impacted by anonymization, which adds to the growing evidence of utility-preserving anonymization techniques.

Keywords. k-anonymity, privacy preserving technique, health data sharing, anonymization, de-identification

1. Introduction

The secondary use of health data requires responsible handling of personal data as defined by the European General Data Protection Regulation (GDPR). Through anonymization techniques, data can be altered in a way that it cannot be related to a person anymore. This is performed via so-called transformation models e.g. suppression

[1] Corresponding Author: Lisa Pilgram, Department of Nephrology and Medical Intensive Care, Charité - Universitätsmedizin Berlin; e-mail: lisa.pilgram@charite.de.

and generalization of values. However, this manipulation can lead to utility loss, which needs to be traded off against the degree of protection achieved. Privacy models such as k-anonymity can exhibit good performance in both, privacy and general-purpose utility measures [1,2]. However, the degree to which general-purpose utility measures correlate with actual utility is less frequently reported.

In our study, we provide a comprehensive evaluation of differently anonymized datasets and evaluate their utility in a concrete application example based on the German Chronic Kidney Disease (GCKD) study.

2. Methods

2.1. Data & anonymization

With more than 10 years of data collection and over 50 peer-reviewed publication GCKD represented a granular and validated dataset with the typical diversity and complexity of a medical dataset. It was composed of 5217 records and 69 variables. We performed qualitative and semi-quantitative risk assessment as proposed by Malin et al. [3] and identified 5 variables that could be used for re-identification, so-called quasi-identifiers (QI), namely age, gender, weight, body-mass index (BMI) and renal biopsy.

Anonymization was realized via full-domain generalization and suppression using ARX with its implemented globally-optimal search algorithm [1]. We defined generalization hierarchies in alignment with the use case. We chose 11-anonymity as conservative and (11,2)-strict-average risk as moderate privacy model [4,5]. Both models put restrictions on the uniqueness of records regarding the QIs, where k-anonymity focuses on the highest uniqueness and strict-average-risk on average uniqueness. Thus, the minimum class size defined by k (11 and 2 respectively) can be translated to the maximum attacker risk (9.09% and 50.0%). In (11,2)-strict-average risk, k' (11) describes the average group size leading to an average risk of 9.09%.

2.2. Evaluation framework

General-purpose utility was obtained using the granularity and discernibility metrics [1]. We further evaluated use case specific utility by replicating already published research results which identified the risk profile of CKD patients [6]. Exemplary analyses were performed on the original and the two anonymized datasets. We calculated proportional or mean 95% confidence intervals (CI) with the Wilson score interval or t-test respectively and reported the overlap in 95% CI lengths as measure of replicability [7]. Age, weight and BMI experienced scale transformation during the anonymization process and were compared visually.

2.3. Ethical statement

All local ethics committees approved the GCKD study (ethics committee Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, no 3831). It is registered in the national registry of clinical studies (DRKS 00003971).

3. Results

3.1. General-purpose utility

(11,2)-strict-average risk exhibited better general-purpose utility: 84.9% granularity (versus 77.3% in 11-anonymity) and 91.0% discernibility (versus 88.6% in 11-anonymity). The same tendency was observed when investigating granularity on QI-level. In QI with scale transformation during the anonymization process (age, weight, BMI), granularity was affected the most ranging between 76.7% and 85.2% in (11,2)-strict-average risk and between 63.2% and 77.6% in 11-anonymity.

3.2. Use case specific utility

We analyzed disease burden stratified by gender and the presence of diabetes mellitus. Results consist of 200 proportion and mean estimates from 40 variables (excluding variables with scale transformation) in 5 subsets. Across all 200 estimates, 95% CI lengths of the original and anonymized data overlapped. Table 1 shows the effects on selected variables of the subset of female non-diabetics. The 95% CI length overlap was lowest in renal biopsy, eGFR and systolic blood pressure. It averaged 90.4% in (11,2)-strict-average risk and 90.2% in 11-anonymity in this subset. For the overall analysis including male diabetics, male non-diabetics, female diabetics, female non-diabetics and the total cohort, the average 95% CI length overlap reached 91.1% for (11,2)-strict-average risk and 89.1% for 11-anonymity. 23.5% (47/200) of the reported estimates were not affected by anonymization at all (95% CI length overlaps 100%). These estimates were (apart from creatinine and cystatin) in the subset of the total cohort where without any stratification only renal biopsy as QI was altered during anonymization.

Table 1. Baseline characteristics of the subset of female non-diabetics within the GCKD cohort across the different datasets. Bpm: beats per minute; (S)(D)BP: (systolic)(diastolic) blood pressure; eGFR: estimated glomerular filtration rate; ARB: angiotensin receptor blockers; n: counts; SD: standard deviation; CI: confidence interval.

	Original dataset			11-anonymity	(11,2)-strict-average risk
	n or mean	SD or %	95% CI	95% CI overlap	95% CI overlap
Family history: stroke	552	39.8	37.2-42.4	84.2	96.4
Family history: hypertension	1013	78.2	75.8-80.4	87.7	88.5
Family history: diabetes	662	48.3	45.6-50.9	59.6	74.6
Family history: renal disease	444	32.9	30.4-35.5	92.6	95.3
Family history: dialysis	110	8.2	6.8-9.9	88.9	95.3
Current smokers	217	15.0	13.2-16.9	93.6	92.2
Former smokers	427	29.4	27.1-31.9	97.1	96.0
Never smokers	807	55.6	53.0-58.2	96.4	91.7
Hospitalization past 5 years	1011	70.6	68.1-72.9	93.0	97.1
Pulse (bpm)	71.3	11.5	70.7-71.9	75.0	72.1
SBP (mmHg)	135.3	19.7	134.0-136.0	50.0	50.0
DBP (mmHg)	80.6	11.0	80.1-81.2	87.1	87.1
BP < 130/80 mmHg	454	31.8	29.4-34.2	78.9	89.0
BP < 140/90 mmHg	863	60.4	57.8-62.9	70.5	72.4

Serum creatinine (mg/dl)	1.3	0.4	1.2-1.3	100.0	100.0
eGFR (ml/min)	49.3	18.4	48.4-50.3	32.5	43.3
eGFR >= 60 ml/min	292	20.2	18.2-22.3	52.4	55.4
eGFR 45-59 ml/min	474	32.7	30.3-35.2	79.3	81.3
eGFR 30-44 ml/min	544	37.5	35.1-40.1	79	81.6
eGFR < 30 ml/min	139	9.6	8.2-11.2	93.6	96.9
Medication: ACE-inhibitors	553	37.8	35.3-40.4	95.3	98.1
Medication ARB	606	41.5	38.9-44.0	96.4	97.2
Medication: diuretics	659	45.1	42.5-47.7	76.7	89.8
Medication: beta blockers	640	43.8	41.2-46.4	78.6	78.6
Renal biopsy	468	32.0	29.7-34.5	24.5	40.8

In a second analysis, we replicated the published age distribution of patients stratified by gender to include an example of scale transformation. Age was given in 5- or 10-years intervals in our anonymized datasets which met the originally reported distribution in 10-years intervals. As demonstrated in Figure 1, results were similar in both anonymized datasets.

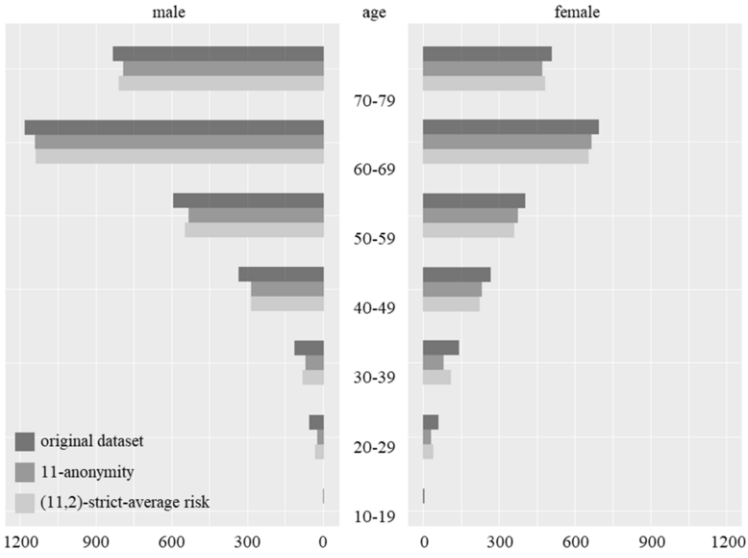


Figure 1. Age distribution of patients enrolled into the GCKD study stratified by gender. Results of the original dataset are illustrated in dark gray followed by the ones of anonymized data using 11-anonymity (middle gray) and (11,2)-strict-average risk (light gray).

4. Discussion

We detected similar tendencies among general-purpose and use case specific utility measures. In both, the moderately strong anonymization approach performed better. Proportion of renal biopsy was impacted most by anonymization in terms of lowest 95% CI length overlaps which might be due to its low case numbers in the individual strata and the relatively high application of suppression to comply to the privacy models.

However, the performance differences between the anonymization approaches seem not to relevantly effect real-world analyses in our scenario. In the exemplary analyses, 95% CI length overlapped, which we see as a confirmation of utility-preserving anonymization achieved by both approaches. The actual choice of the privacy model should consider the exact data sharing scenario (e.g. controlled vs. open access) and further safeguards implemented.

Our results are in line with other real-world applications. In the Lean European Open Survey on SARS-CoV-2 infected patients, for example, anonymization only led to deviations of not more than 0.11% in reported frequencies [8]. Song et al. could replicate prediction accuracy of a machine learning model for early acute kidney injury risk prediction in anonymized data [9]. Such more complex correlations need to be addressed more often in further research. Even though descriptive statistics as we did in our study represent an almost mandatory part of any study, anonymized data must also be able to reproduce clinical implications drawn from more sophisticated analyses.

5. Conclusion

The presented evaluation contributes to the growing evidence of successful utility-preserving anonymization techniques. This evidence is needed to address ambiguities about the trade-off between privacy and utility, and to foster adoption of anonymization into standard data management procedures.

References

[1] Prasser F, Eicher J, Spengler H, Bild R, Kuhn KA. Flexible Data Anonymization Using ARX — Current Status and Challenges Ahead. J Software Pract Exper 50, 7 (2020);1277-1304. doi: 10.1002/spe

[2] Haber AC, Sax U, Prasser F; NFDI4Health Consortium. Open tools for quantitative anonymization of tabular phenotype data: literature review. Brief Bioinform. 2022 Nov 19;23(6):bbac440. doi: 10.1093/bib/bbac440. PMID: 36215114; PMCID: PMC9677485.

[3] Malin B, Loukides G, Benitez K, Clayton EW. Identifiability in biobanks: models, measures, and mitigation strategies. Hum Genet. 2011 Sep;130(3):383-92. doi: 10.1007/s00439-011-1042-5. Epub 2011 Jul 8. PMID: 21739176; PMCID: PMC3621020.

[4] Sweeney L. K-Anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10.5 (2002): 557-570. doi: 10.1142/s0218488502001648

[5] Committee on Strategies for Responsible Sharing of Clinical Trial Data; Board on Health Sciences Policy; Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington (DC): National Academies Press (US); 2015 Apr 20. doi: 10.17226/18998. PMID: 25590113.

[6] Titze S, Schmid M, Köttgen A, Busch M, Floege J, Wanner C, Kronenberg F, Eckardt KU; GCKD study investigators. Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort. Nephrol Dial Transplant. 2015 Mar;30(3):441-51. doi: 10.1093/ndt/gfu294. Epub 2014 Sep 30. PMID: 25271006.

[7] Karr AF, Kohnen CN., Oganian A, Reiter JP, Samil AP. A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality. The American Statistician 2006, 60:3:224-232. doi: 10.1198/000313006X124640

[8] Jakob CEM, Kohlmayer F, Meurers T, Vehreschild JJ, Prasser F. Design and evaluation of a data anonymization pipeline to promote Open Science on COVID-19. Sci Data. 2020 Dec 10;7(1):435. doi: 10.1038/s41597-020-00773-y. PMID: 33303746; PMCID: PMC7729909.

[9] Song X, Waitman LR, Hu Y, Luo B, Li F, Liu M. The Impact of Medical Big Data Anonymization on Early Acute Kidney Injury Risk Prediction. AMIA Jt Summits Transl Sci Proc. 2020 May 30;2020:617-625. PMID: 32477684; PMCID: PMC7233037.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230059

The Necessity of Multiple Data Sources for ECG-Based Machine Learning Models

Lucas PLAGWITZ[a,1], Tobias VOGELSANG[a], Florian DOLDI[b], Lucas BICKMANN[a], Michael FUJARSKI[a], Lars ECKARDT[b], and Julian VARGHESE[a]
[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department for Cardiology II-Electrophysiology, University Hospital Münster, Germany*

Abstract. Even though the interest in machine learning studies is growing significantly, especially in medicine, the imbalance between study results and clinical relevance is more pronounced than ever. The reasons for this include data quality and interoperability issues. Hence, we aimed at examining site- and study-specific differences in publicly available standard electrocardiogram (ECG) datasets, which in theory should be interoperable by consistent 12-lead definition, sampling rate, and measurement duration. The focus lies upon the question of whether even slight study peculiarities can affect the stability of trained machine learning models. To this end, the performances of modern network architectures as well as unsupervised pattern detection algorithms are investigated across different datasets. Overall, this is intended to examine the generalization of machine learning results of single-site ECG studies.

Keywords. data integration, ECG, machine learning, external validation

1. Introduction

The research trend on decision support systems via machine learning (ML) continues unabated in many disciplines. However, analysis of ML algorithms and pattern recognition for medical problems is subject to strong bias, as it consists mainly of retrospective data that are insufficient for robust clinical application and cannot adequately measure the underlying phenomenon, since they usually consider only one data source. This can be especially problematic when a data collection is not standardized, as in the case of magnetic resonance imaging, where measurements are dependent on the device and sequence [1]. A transfer of trained decision models to other datasets is therefore hardly possible. However, even with standardized data acquisition, which is given in the case of standard electrocardiograms (ECGs) with 12-channel array, 10 seconds duration, and a sampling frequency of 500 HZ, different devices and preprocessing steps potentially alter the outcome. The question arises whether this is sufficient to cause an impact on a trained model. Preliminary work already shows the broad-based data collection and interoperability problems at the level of ECG hardware,

[1] Corresponding Author: Lucas Plagwitz, E-mail: lucas.plagwitz@uni-muenster.de.

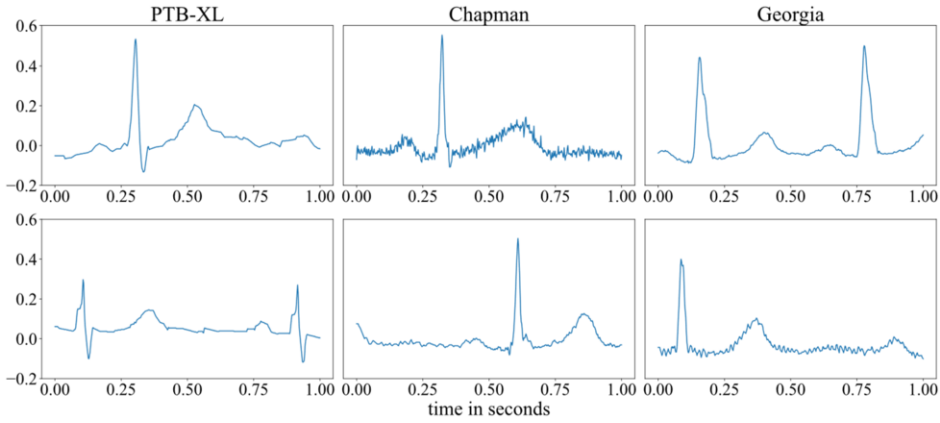


Figure 1. Overview of different ECG measurements depending on the data source. Each column contains two scaled sinus rhythm recordings in lead II of the corresponding data sources (table header).

software, and file formats [2, 3]. Despite this, an in-depth analysis of dataset-specific signal differences is still lacking. To this end, we consider three publicly available data sources that contain quite different ECGs, as shown in Figure 1. Based on these data, we examine the transferability and vulnerability of ML algorithms between these datasets in supervised and unsupervised learning settings.

2. Methods

2.1. Datasets

To investigate comparability between different ECG studies, we investigated the three largest freely available clinical 12-lead ECG datasets (10s length and sampling frequency of 500 HZ) hosted on the PhysioNet online database at the time of this study:

- The PTB-XL dataset contains 21799 ECGs from 18869 patients recorded between years 1989-1996 with Schiller AG equipment [4]. In addition to the raw time signals, information on diagnosis, shape, and rhythm is available.
- Second, a large-scale arrhythmia database from Chapman University is considered, providing ECGs from 45152 patients [5]. These data were collected and stored using devices from General Electric (GE).
- The third dataset, called Georgia, was collected through Emory University, Atlanta, Georgia, and represents a large population of 10344 patients in the southeastern United States. The least is known about this dataset, except that it was provided by Emory University via PhysioNet as part of a 2020 ECG classification challenge [6].

Since the three datasets were not scaled consistently, we applied sample-based abs-max scaling: The entire 12-lead ECG was divided by its highest absolute value. This results in a loss of important information, therefore the performance results are no longer comparable with other studies, but it still enabled us to exclude scaling effects between different data sources.

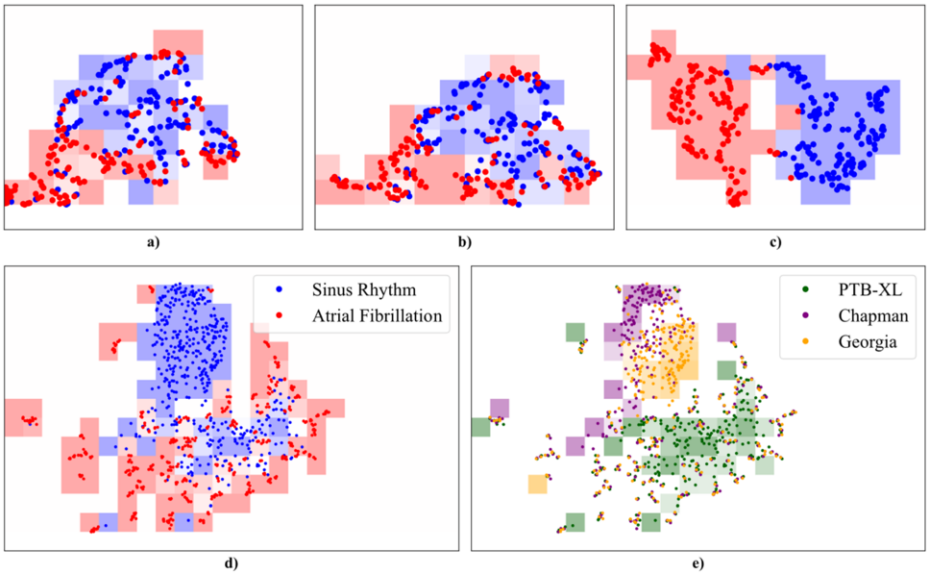


Figure 2. Results of clustering analysis reveal source-specific differences. a) Separation results for PTB-XL samples, labeled according to SR (blue) and AF (red). b) Same analytical procedure as a) based on the Chapman dataset. c) Result of the separation analysis for samples from the Georgia dataset. d) Clustering analysis performed on all datasets simultaneously labeled according to SR and AF. e) Same distribution as in d) relabeled according to the data source. The color of the squares indicates the distribution of the labels within this space.

2.2. Unsupervised analysis

To visualize the influence of possible dataset-specific patterns, we considered two diagnostic subclasses of all datasets: For each dataset, 150 ECGs with labeled atrial fibrillation (AF) and 150 sinus rhythm (SR) ECGs were selected. Measurements were matched in two ways: between diagnoses in a dataset and between data sources in terms of age and sex. Four different tests with the same procedure were calculated based on different data bases: three times for the individual sets and once for all datasets. To compare the scaled ECG data in a meaningful way, we applied a word representation based on the Bag of Symbolic Fourier Approximation Symbols (BOSS) method [7]. This representation was then applied as input to the Uniform Manifold Approximation and Projection (UMAP) algorithm to create a two-dimensional visualization of the data [8].

2.3. Supervised learning

To investigate the model validity of predictions across dataset boundaries, we systematically trained models with a stratified 5-fold cross-validation on one of the datasets, followed by a test of this model on the other two sets. Four different binary classifications were considered: sex, age (>50), AF vs. SR, and first-degree AV block (1AVB) vs. SR. These are all matched as best as possible for age and sex. For the train and test procedure we utilized two convolutional neural networks (CNNs) from PyTorch and tsai projects: a fully convolutional network (FCN) and XceptionTime (XcTime) architecture [9]. The performances were compared using the balanced accuracy score (BACC) based on an iterative scheme whereby training was performed three times on each set and the trained model is additionally tested on the remaining two data sources (known: 3, reference: 6).

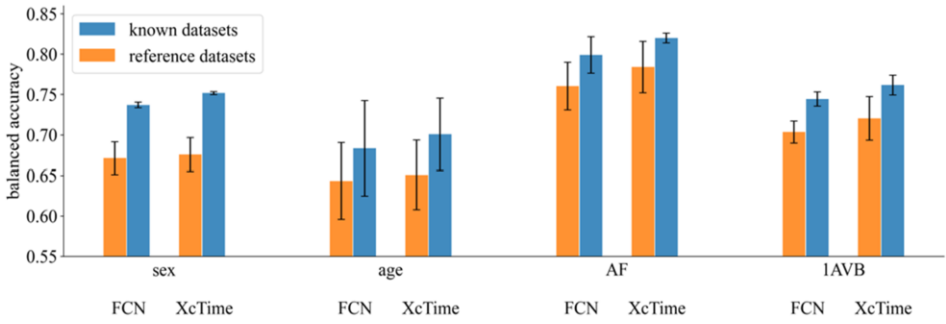


Figure 3. Transferability of classification models across datasets per architecture and label. The blue bar represents the average three test performances based on the known dataset. The orange bar illustrates the average six test performances based on the reference datasets, which source differs from the training set.

3. Results

Figure 2 summarizes the results of the cluster analysis. It was obtainable in the first three subfigures that AF separated well from SR ECGs. Especially on the Georgia dataset, the two labels split almost homogeneously, while PTB-XL and Chapman still had AF measurements within SR regions. In addition to these plots, the cluster distribution was calculated across the three datasets using the same procedure. On the one hand, the AF-SR dependent areas were still recognizable to some extent, but on the other hand, there were separable regions due to the source.

Table 1. List of all investigated ECG datasets with their class distribution per dataset. The matching was done across all data sources.

	sex	**age (> 50)**	**AF**	**1AVB**
dataset distribution	793 / 700	948 / 552	700 / 568	679 / 277
subset	SR (age matching)	SR (sex matching)	SR + AF (age, sex matching)	SR + 1AVB (age, sex matching)

For every label, a sub-dataset was created based on different subsets and properties. Table 1 lists the underlying criteria and class distribution per data source. Two CNN architectures were trained on each of these problems. Figure 3 presents the average model performance based on the unknown and familiar datasets. First, the performances differed between classification label between age (known: 0.7, reference: 0.65) to AF (known: 0.8, reference: 0.76). Furthermore, it was noticeable on every instance that the performance for the known datasets was consistently better compared to the non-trained reference datasets. The difference of known and reference performance varied depending on the label. For example, in the case of sex recognition, the difference in BACC was 0.08, whereas it was 0.04 in the case of AF and 1AVB. Besides this effect, a continuously higher BACC was shown for the XceptionTime model compared to the FCN by about 0.025.

4. Discussion and Conclusion

The results of our experiments showed a clear impact of dataset-specific features on ML algorithms. While a diagnosis-based characterization clustering was observable, delineations depending on the data source clearly emerged. The distinction of AF and SR was possible, but minor diagnosis-based effect sizes could be completely obscured by such artifacts. However, unsupervised methods on ECG data account for only a small fraction of ML applications. According to our research, supervised approaches were influenced by source-specific characteristics as well. We have shown that for four different binary classification tasks, transferability of the model to other datasets was associated with lower predictive performance. This varied between 0.04 - 0.08 BACC depending on the label and architecture. We used a matching procedure to align demographic characteristics between datasets as much as possible. However, annotation- or cohort-specific characteristics (apart from age and sex) could not be eliminated.

Because this effect has significant implications for the application of ML studies to clinical practice, we advocate observing study-specific effects (such as device type or preprocessing steps) when constructing a predictive model. As a by-product, comparable to previous work, we endorse more complex CNN models when it comes to pure predictive performance - even on relatively small sample sizes [10].

Overall, the impact of dataset-specific characteristics on ML algorithms was evaluated for ECG data. Both unsupervised and supervised analyses were revealed to be affected by these side effects. To get one step closer to the underlying medical phenomenon and thus to clinical relevance, we plead for increasingly frequent external validation of study results or models that were already trained across multiple data sources.

References

[1] Biberacher V, Schmidt P, Keshavan A, Boucard C, Righart R, Sämann P, Preibisch C, Fröbel D, Aly L, Hemmer B, Zimmer C, Henry RG, Mühlau M. Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. Neuroimage. 2016; 142.

[2] Husain K, Zahid MSM, Hassan SU, Hasbullah S, Mandala S. Advances of ECG Sensors from Hardware, Software and Format Interoperability Perspectives. Electronics. 2021; 10, 105.

[3] Stamenov D, Gusev M, Armenski G. Interoperability of ECG standards. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2018.

[4] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL: A Large Publicly Available ECG Dataset. Scientific Data. 2020: 7, 154.

[5] Zheng J, Guo H, Chu H. A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0.0). PhysioNet. 2022.

[6] Perez Alday EA, Gu A, Shah A, Liu C, Sharma A, Seyedi S, Bahrami Rad A, Reyna M, Clifford G. Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020 (version 1.0.2). PhysioNet. 2022.

[7] Schäfer P. The BOSS is concerned with time series classification in the presence of noise. Data Mining Knowledge. 2015; 29.

[8] McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint. 2018 arXiv:1802.03426.

[9] Oguiza I. tsai - A state-of-the-art deep learning library for time series and sequential data. GitHub. 2022; https://github.com/timeseriesAI/tsai.

[10] Doldi F, Plagwitz L, Hoffmann LP, Rath B, Frommeyer G, Reinke F, Leitz P, Büscher A, Güner F, Brix T, Wegner FK, Willy K, Hanel Y, Dittmann S, Haverkamp W, Schulze-Bahr E, Varghese J, Eckardt L. Detection of Patients with Congenital and Often Concealed Long-QT Syndrome by Novel Deep Learning Models. Journal of Personalized Medicine. 2022; 12, 1135.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230060

Predicting Progression of Type 2 Diabetes Using Primary Care Data with the Help of Machine Learning

Berk OZTURK[a,1], Tom LAWTON[a,b], Stephen SMITH[a], Ibrahim HABLI[a]
[a] *University of York, York, YO10 5GH, UK*
[b] *Bradford Teaching Hospitals NHS Foundation Trust, Bradford BD9 6RJ, UK*

Abstract. Type 2 diabetes is a life-long health condition, and as it progresses, a range of comorbidities can develop. The prevalence of diabetes has increased gradually, and it is expected that 642 million adults will be living with diabetes by 2040. Early and proper interventions for managing diabetes-related comorbidities are important. In this study, we propose a Machine Learning (ML) model for predicting the risk of developing hypertension for patients who already have Type 2 diabetes. We used the Connected Bradford dataset, consisting of 1.4 million patients, as our main dataset for data analysis and model building. As a result of data analysis, we found that hypertension is the most frequent observation among patients having Type 2 diabetes. Since hypertension is very important to predict clinically poor outcomes such as risk of heart, brain, kidney, and other diseases, it is crucial to make early and accurate predictions of the risk of having hypertension for Type 2 diabetic patients. We used Naïve Bayes (NB), Neural Network (NN), Random Forest (RF), and Support Vector Machine (SVM) to train our model. Then we ensembled these models to see the potential performance improvement. The ensemble method gave the best classification performance values of accuracy and kappa values of 0.9525 and 0.2183, respectively. We concluded that predicting the risk of developing hypertension for Type 2 diabetic patients using ML provides a promising stepping stone for preventing the Type 2 diabetes progression.

Keywords. Type 2 diabetes, comorbidity, machine learning, healthcare, data quality

1. Introduction

Diabetes is one of the most common health conditions in the world, and its incidence in the population has been increasing rapidly over the years [1]. It is a lifelong health condition that occurs when the pancreas cannot produce enough insulin to balance blood sugar (blood glucose) [2]. This health condition can be observed at any age and, if not managed properly, can progress and develop comorbidities [3]. Progression of these comorbidities could result in a range of poor outcomes for the patient, including blindness, the need for dialysis, heart attacks and strokes, the need for limb amputation, and even mortality [4].

[1] Corresponding Author: Berk Ozturk, Tel.: +44 751 337 5115.
E-mail addresses: berk.ozturk@york.ac.uk (B. Ozturk), tom.lawton@bthft.nhs.uk (T. Lawton), stephen.smith@york.ac.uk (S. Smith), ibrahim.habli@york.ac.uk (I. Habli).

Early diagnosis of diabetes progression and proper management of diabetes can significantly prevent the progression of diabetes [5]. Further, it is well known that hypertension is one of the most important risk factors for developing comorbidities [6]. Because diabetes and hypertension are synergistically dangerous, early and accurate prediction of hypertension risk is crucial in terms of managing the progression of diabetes.

Type 1 and Type 2 diabetes are the two main types of diabetes in the world. Type 1 diabetes occurs when the body's immune system targets and destroys the insulin-producing cells in the pancreas [7]. This can develop quickly and requires regular insulin injections [3]. Type 2 diabetes develops when pancreatic cells are not able to produce enough insulin or when body cells do not react to insulin [7]. People may have Type 2 diabetes without realising it because the early symptoms tend to be ambiguous [4]. However, this type of diabetes can be managed by changing lifestyle and employing appropriate treatment methods for the comorbidities [4]. Type 2 is the most common diabetes type with 90% of incidence in the world population [8]. Therefore, we focused to develop our model for Type 2 diabetic patients.

The majority of Machine Learning (ML)-based Type 2 management studies in the literature are related to the early diagnosis of Type 2 Diabetes, and it is noteworthy that there has not been a sufficient amount of work on prediction of Type 2 diabetes-related comorbidities [9]. Therefore, studies to predict the risk of developing comorbidities in Type 2 diabetes patients and to manage this health condition have great importance. This makes it more critical for Type 2 diabetic patients with hypertension because Type 2 diabetes combined with hypertension increases the likelihood of severe comorbidities. This study aims to predict the risk of developing hypertension and reduce the risk of developing further critical comorbidities in Type 2 patients to support lifelong Type 2 diabetes management.

2. Methods

2.1. Data Preprocessing

One of the most important elements in ML-based Type 2 diabetes problems is the real-life representation of the data used. In this study, Primary Care data of patients in the Connected Bradford dataset were used. Connected Bradford Primary Care is a large dataset containing all the observation data of primary care healthcare institutions in Bradford, UK. In this dataset, patients were assigned an anonymous unique ID, and each of the observations resulting from their visits to these primary care health institutions was recorded. The dataset used in this study consists of 1,058,139 patient entries (rows) and each entry has feature (variable) columns such as person id, observation definition, numeric laboratory result, observation date, etc.

However, since this study is a Type 2-related study, these data were filtered only with patients who already had Type 2 diabetes. After this filtering process, a total of over 476,000 Type 2 diabetic patients with over 14,000 variable columns remained. Since this dataset is still very large and calculations are time-consuming, a randomly generated subset of one million rows with 43,000 patients was used. In parallel, the observation definition column was grouped by diseases, and it has been observed that the most common observation is hypertension. To this end, hypertension has been identified as a marker of the risk of Type 2 diabetes progression.

In this study, in accordance with the National Institute for Health and Care Excellence (NICE), patients with 7-day mean systolic and diastolic blood pressure readings of more than 140 mmHg and 90 mmHg, respectively, were categorised as at high risk of developing hypertension [4]. The patients with lower than these reading values were categorised as having a low risk of developing hypertension. Since the finalised data frame is still very big, the columns with no data have been eliminated, and the most 20 frequent variables in this database have been selected to ensure computation and time efficiency. Since there are still some missing values in the data, these missing values were imputed with the average value for each column. Then the resulting new data frame is scaled.

2.2. ML Methods

The most common ML algorithms for Type 2 diabetes-related problems were used, namely [10]: Naïve Bayes (NB), Neural Network (NN), Random Forest (RF), and Support Vector Machine (SVM). These four ML algorithms were trained separately. 80% of the data have been trained using k-Fold cross-validation and parameter hyper-tuning. Cross-validation was used to reduce overfitting and prevent bias. Parameter hyper-tuning was used to optimise the performance of trained models. The remaining 20% of the data was used for testing. Next, Generalized Linear Model (GLM) was used as an ensemble method to combine the predictions and increase the prediction performance of our classification problem.

3. Results

Table 1 shows the Accuracy and Kappa values of the ML methods used. These values are the default metrics used to evaluate algorithms on binary and multi-class classification datasets with the Caret package in R. Accuracy is the percentage of correctly classified instances out of all instances [11]. Kappa or Cohen's Kappa is a classification accuracy, except that it is normalised at the baseline of random chance on the dataset [11]. It is a more useful measure for problems that have an imbalance in the classes [12].

Table 1. Performance values of each ML method and ensemble method

ML Method	Accuracy	Kappa
Naïve Bayes	0.8667	0.1296
Neural Network	0.9500	-0.0034
Random Forest	0.9504	0.0000
Support Vector Machine	0.9500	-0.0008
Ensemble	0.9525	0.2183

Figure 1 shows the importance of each variable in the preprocessed dataset. These importance values have been calculated for the ensemble method using the Caret package in R. In this subset, Body Mass Index – Observation (i.e. body mass divided by the square of the body height), Neutrophil Count (i.e. subset of white blood cells in the immune system), and Serum Cholesterol Level (i.e. combined amount of bad and good cholesterol in blood), are the three most important variables in predicting the risk of hypertension in a patient with Type 2, respectively.

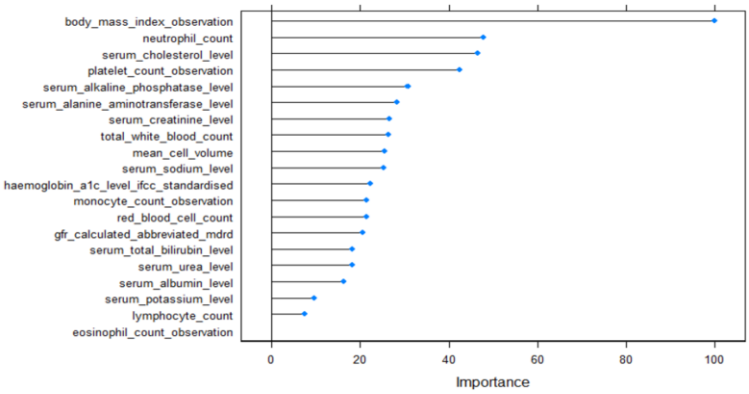


Figure 1. Importance level of each feature in the ensemble method

4. Discussion

In Table 1, RF has the highest accuracy in predicting the risk of hypertension, but the Kappa value, another important performance criterion in classification problems, is 0.0000. Despite its lower accuracy, NB has the highest Kappa value among the four ML algorithms with 0.1296. This shows that accuracy and kappa may give different results with repeated use of ML algorithms, and the performance of the model may vary according to the intended performance criteria. However, when the accuracy and kappa values of the ensemble method are considered, it is seen that both performance criteria have the highest values with 0.9525 and 0.2183, respectively. Even though the accuracy does not increase significantly, the increase in the Kappa value is noteworthy.

In Figure 1, it is observed that the prediction of risk of hypertension in patients with Type 2 diabetes may depend on various variables. However, it has been observed that the importance of variables may vary and especially some variables may have greater importance in the ML model. This causes ML-based prediction outcomes to be more affected by some variables. Seeing the importance levels of the variables in the ML-based models provides an opportunity to make interpretations of the reasons behind the outcomes of the ML model.

5. Conclusion

Hypertension is the most common comorbidity among Type 2 diabetic patients in the Connected Bradford data. In addition, hypertension is related to many serious diseases and has a great significance in predicting the clinically important poor outcomes in Type 2 diabetes, such as heart attack, blindness, or neurological issues. It is important to predict and manage hypertension, which is one of the most important risk factors for developing serious comorbidities in Type 2 diabetic patients. We showed that when high-performing ML models are ensembled, the performance values of the final prediction can potentially increase. However, it is noteworthy to criticise that using different ML algorithms with large datasets has limitations in prediction of Type 2 diabetes progression. Since there are no distinct ML-based methods in progression of Type 2 diabetes, it is crucial to choose the most related data pre-processing techniques and ML

algorithms. It also is useful to provide feature importance of the trained ML model to see the impact of each variable on prediction of Type 2 diabetes progression. Finally, we aim to continue to improve the performance and robustness of the ML algorithms and importantly develop a clinical safety case that considers the assurance of the algorithms in the intended clinical workflow and setting [13]. Beyond performance and safety, we plan to consider the wider ethical issues with the deployment of this type of clinical ML algorithms and explore questions of legal liability and moral responsibility for the outcome of the ML-based decision support systems [14].

Acknowledgment

This work was supported by the Assuring Autonomy International Programme, a partnership between Lloyd's Register Foundation and the University of York and by the Engineering and Physical Sciences Research Council (EP/W011239/1).

References

[1] Farran B, Channanath AM, Behbehani K, Thanaraj TA. Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study. BMJ open. 2013 Jan 1;3(5):e002457. doi: 10.1136/bmjopen-2012002457

[2] K, Bunescu R, Marling C, Shubrook J, Schwartz F. A machine learning approach to predicting blood glucose levels for diabetes management. InWorkshops at the Twenty-Eighth AAAI conference on artificial intelligence 2014 Jun 18.

[3] Peddinti G, Cobb J, Yengo L, Froguel P, Kravić J, Balkau B, Tuomi T, Aittokallio T, Groop L. Early metabolic markers identify potential targets for the prevention of type 2 diabetes. Diabetologia. 2017 Sep;60(9):1740-50. doi: 10.1007/s00125-017-4325-0

[4] Alonso-Morán E, Orueta JF, Esteban JI, Axpe JM, González M, Polanco NT, Loiola PE, Gaztambide S, Nuño-Solinis R. The prevalence of diabetes-related complications and multimorbidity in the population with type 2 diabetes mellitus in the Basque Country. BMC public health. 2014 Dec;14(1):1-9.

[5] Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Scientific reports. 2020 Jul 20;10(1):1-2.

[6] Long AN, Dagogo-Jack S. Comorbidities of diabetes and hypertension: mechanisms and approach to target organ protection. The journal of clinical hypertension. 2011 Apr;13(4):244-51.

[7] Ozougwu JC, Obimba KC, Belonwu CD, Unakalamba CB. The pathogenesis and pathophysiology of type 1 and type 2 diabetes mellitus. J Physiol Pathophysiol. 2013 Sep 30;4(4):46-57.

[8] DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, Hu FB, Kahn CR, Raz I, Shulman GI, Simonson DC. Type 2 diabetes mellitus. Nature reviews Disease primers. 2015 Jul 23;1(1):1-22.

[9] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal. 2017 Jan 1;15:104-16. doi: 10.1016/j.csbj.2016.12.005

[10] Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. Journal of medical Internet research. 2018 May 30;20(5):e10775. doi: 10.2196/10775.

[11] Kuhn M. Building predictive models in R using the caret package. Journal of statistical software. 2008 Nov 10;28:1-26.

[12] Foody GM. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. Remote Sensing of Environment. 2020 Mar 15;239:111630.

[13] Hawkins R, Paterson C, Picardi C, Jia Y, Calinescu R, Habli I. Guidance on the assurance of machine learning in autonomous systems (AMLAS). arXiv preprint arXiv:2102.01564. 2021 Feb 2.

[14] Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bulletin of the World Health Organization. 2020 Apr 4;98(4):251.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230061

Breaking Barriers for Interoperability: A Reference Implementation of CSV-FHIR Transformation Using Open-Source Tools

Samer ALKARKOUKLY[a,b,1], Md. Mostafa KAMAL[a,b] and Oya BEYAN[a,c]
[a] *Institute for Biomedical Informatics, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany*
[b] *CECAD, University of Cologne and University Hospital Cologne, Cologne, Germany*
[c] *Department of Data Science and Artificial Intelligence, Fraunhofer FIT, Germany*
ORCiD ID: Samer Alkarkoukly https://orcid.org/0000-0003-1891-1366, Md. Mostafa Kamal https://orcid.org/0000-0002-5506-4875, Oya Beyan https://orcid.org/0000-0001-7611-3501

Abstract. FHIR is a widely accepted interoperability standard for exchanging medical data, but data transformation from the primary health information systems into FHIR is usually challenging and requires advanced technical skills and infrastructure. There is a critical need for low-cost solutions, and using Mirth Connect as an open-source tool provides this opportunity. We developed a reference implementation to transform data from CSV (the most common data format) into FHIR resources using Mirth Connect without any advanced technical resources or programming skills. This reference implementation is tested successfully for both quality and performance, and it enables reproducing and improving the implemented approach by healthcare providers to transform raw data into FHIR resources. For ensuring replicability, the used channel, mapping, and templates are available publicly on GitHub[2].

Keywords. FHIR, Fast Health Interoperability Resources, Interoperability, Data Transformation, CSV, Mirth Connect, open-source

1. Introduction

Transforming raw data from healthcare information systems into FHIR resources usually requires advanced technical skills and software. The aim of this work is to propose a user-friendly method to transform medical data from CSV format into FHIR resources without requiring sophisticated programming scripts or tools. It uses an open-source tool and publicly available transformation "channels" and templates, and could be reproduced and improved by any healthcare provider with basic knowledge and skills.

Healthcare providers collect different types of patient data and store them in Electronic Medical Records "EMRs" which are defined as computerised medical information systems that collect, store, and display patient information [1]. EMRs in healthcare institutions reflect the differences among the several information systems used

[1] Corresponding Author: Samer Alkarkoukly, E-mail: Mohammad.abboud-alkarkoukly@uk-koeln.de
[2] https://github.com/alkarkoukly/CSV-FHIR-Transformer

in the different departments of each healthcare facility, which lead the medical data to be fragmented and spread over different systems. Data fragmentation is one of the challenges against the interoperability of EMRs. To enable the secondary usage of clinical data, several health information systems providers permit data export in Comma Separated Values (CSV) format, which is the most frequently used data format in storing tabular data as a two-dimensional array. However, CSV does not describe metadata such as datatype, data quality, or data provenance [2], and CSV files fail to provide semantic interoperability with other systems or other potential users of the data.

In order to facilitate medical data exchange and to enhance interoperability, Fast Health Interoperability Resources (FHIR) were designed and used. FHIR is a standard for exchanging healthcare information electronically [3] and it maintains syntactic interoperability and ensures correct interpretation of the data at the receiver end. Also, it maintains semantic interoperability by using terminologies and coding systems. Additionally, transferring data in FHIR format enhances the FAIR principles (Findable, Accessible, Interoperable, Reusable) which guide scientific data management and stewardship, and are relevant to all stakeholders in the digital health ecosystem [4].

To enhance the reusability and interoperability of clinical data, several tools were designed to receive, filter, clean, transform, export and validate the collected data in the healthcare facilities' primary information systems. One of these open-source tools is Mirth Connect, developed in 2006 as MIRTH, an acronym for "Messaging, Interfaces, Routing, Transformations, Healthcare" [5]. It is an integration engine that reads patient data from different sources (database, local directory, local server and many others), transforms the data and translates messages standards into different ones. Mirth Connect can read and export files from different formats (CSV, JSON, XML, HL7 V2 messages).

The literature review showed that the reference implementation of this work has not been implemented yet and would be an added knowledge to the work in this field. The most relevant published work included converting HL7 ADT messages into FHIR using an open-source tool called XSLT [6], but the approach seems to be focused on oncology data and only on ADT messages. Another article focused on the concept and the logical mapping and the architectural differences between openEHR and FHIR and not detailing the technical steps followed during the transformation [7]. A third approach [8] discussed the concept and challenges against converting data to FHIR using the tool "CAMP FHIR" which can get the data from a relational database and convert to FHIR with no details about the followed steps.

Currently, most of the available applications for transforming CSV data into FHIR require a Java-based integration service, which requires advanced technical skills in Java development. In this work, the transformation was done without coding, in a user-friendly approach and interface. The additional added value of this work, is the usage of the open-source integration engine "Mirth Connect", which is widely used to manage HL7 V2 messages in healthcare facilities, in transforming data into FHIR.

2. Methods

This reference implementation uses Mirth Connect, as an open-source data integration engine, to transform CSV files exported from the healthcare information system, or filled manually during data collection, and saved in a local folder, then exports the data as FHIR Patient resources in so-called "transaction bundle" to two destinations: 1) a FHIR server. 2) a local folder in the same PC as JSON files. Figure 1 illustrates the process:

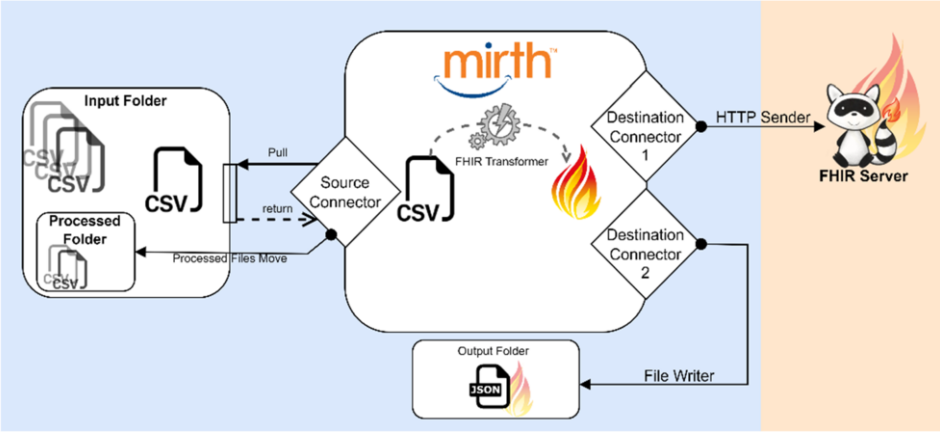


Figure 1. Illustration of Mirth Connect Channel for CSV-FHIR Transformation

Mirth Connect (MC) could be downloaded and installed from GitHub or from the vendor. The installation includes four components MC Server, MC Server Manager, MC Command Line Interface, and MC Administrator Launcher, and it will obtain two ports from the local PC. After installation, all the following work can be done using the graphical user interface of Mirth Connect by running the server manager first, starting MC service, and then launching MC administrator. A channel needs to be designed and configured in MC Administrator; the configuration settings are provided in Table 1.

Table 1. Mirth Connect CSV-FHIR Transformer Channel configuration

Connector	Data Type	Connector Type	Configuration
Source	Delimited Text	File Reader	Directory: Local PC input folder Filename Filter Pattern: "*.csv"
Destination 1	JSON	HTTP Sender	URL: FHIR server endpoint Content-type: "application/JSON" Content: "${JsonUtil.prettyPrint(${message.encodedData})}"
Destination 2	JSON	File Writer	Directory: Local PC output folder File Name: "${originalFilename}${COUNT}.json" Template: Same value of "content" of destination 1

Transformer Configuration: The source connector has a transformer which transforms the CSV inbound message from the source into JSON outbound messages as FHIR resources in a transaction bundle for both of the planned destinations. The inbound message template is a predefined CSV file reflecting the FHIR Patient resource elements and all the incoming messages should use the same template. The template is available publicly in the project's GitHub repository and could be easily extended and improved. The outbound message has a JSON structure obtained from FHIR Patient resource. The mapping between the elements of inbound CSV and outbound JSON is done easily by drag-and-drop from the inbound message tree into the outbound message tree.

After deploying the channel, Mirth Connect automatically reads the entire content of all CSV files in the input folder, transforms the content into FHIR resources, then sends the generated resources into a locally installed FHIR server and at the same time, exports them as JSON files in the output folder. All the processed CSV files will be moved into another folder "Processed". The processes of this reference implementation are just an example of the implementation of CSV-FHIR transformation, and they are illustrated in Figure 1. This scenario proves that the followed methodology could be implemented in any other use-case of raw data transformation to FHIR. The input data

could be any of the different file formats (XML, JSON, HL7, TSV, DICOM) or any of the different types of source connectors (HTTP Listener, Channel Reader, Database Reader, TCP Listener) or even a response from REST interface (such as another FHIR server). To test the channel, 21 synthetic datasets of randomised real personal data were generated in CSV files and used to test the quality and performance of the channel.

3. Results

The suggested reference implementation was tested successfully by exporting FHIR resources to a locally installed HAPI server and to a local PC directory, and additional quality and performance tests were implemented to check the applicability on large-scale implementations and also the validity and correctness of the generated FHIR resources. The performance test included datasets in CSV format from the input folder in 3 scenarios (10 files of 1,000, 10 files of 10,000, 1 file of 50,000 records); the processing speed average was 59 milliseconds per CSV record, and it is correlated with records amount. The quality test included monitoring the errors received from the channel logs, which was limited to only one timeout error during processing the largest CSV file. Additionally, a validation test of the generated FHIR resources was implemented by validating a random sample of 1% of the processed records (700 resources) using the "$validate" function of the destination server, and 0.1% (70 resources) were checked for correctness by comparing the generated resources with the original data. All the tested resources were valid with no warnings or errors, and matched the original data.

During the implementation, none of the programming languages was needed neither in installation (done with the graphical user interface) nor in the channel set-up (done by drag and drop between the inbound and outbound elements). All the used files, including the exported channel, CSV and JSON templates, and the testing CSV files, are available publicly on GitHub https://github.com/alkarkoukly/CSV-FHIR-Transformer.

4. Discussion

Using Mirth Connect as an open-source tool showed full success in transforming patient data from CSV format into FHIR Patient resources represented in transaction bundles. This approach is suitable for all types of healthcare facilities planning to generate FHIR resources from their primary information systems using basic technical skills and IT infrastructure, without the need for advanced hardware or virtual machines, and without any external services built with Java, C, or Python. The approach is applicable on different types of data sources (regular export in CSV, SQL queries, live streams of files or messages, and services with REST APIs). The reference implementation targeted Patient resource from FHIR specifications, and it is applicable on all of the other types of FHIR resources. It is also possible to include data cleaning, files filtering, data values mapping to terminologies or value sets, and variable format transformations in the channel of Mirth Connect. This could be done using the JavaScript programming environment contained in Mirth Connect user interface. Although the input CSV structure is static, it can be configured with the user interface to meet the different needs. The limitation of this approach is mainly related to the communication with the destination FHIR server and interactively exchange data in both ways during the submission of each single resource. For example, the need for checking if a specific

resource is already available in the FHIR server and then to conditionally create or update the intended resource based on specific elements. This situation is still theoretically possible with Mirth Connect by processing the response from the server. However, it requires further research and testing which are not covered in this paper. The approach followed in this work is reproducible, repeatable and improvable by using the exported channel, templates and testing files provided in the project repository in GitHub.

5. Conclusions

Transforming raw data from CSV into FHIR resources usually requires advanced technical and programming skills, and is not easily applicable alone by the healthcare provider or the data scientist. By using Mirth Connect as an open-source and user-friendly mapping and transformation tool, this reference implementation proves the possibility of transforming CSV into FHIR Patient resources and sending them to a FHIR server and a local directory. This work successfully tested the quality and performance of the implemented approach on large-scale implementations.

The channel, templates, and testing files used in this work are available publicly, and could be used to reproduce and improve the implemented approach. Future research on the topic can demonstrate further implementation possibilities with other types of FHIR resources and including advanced filtering, transformation and mapping processes.

Acknowledgment

This work was partially supported by the HiGHmed Consortium funded by the German Federal Ministry of Education and Research (BMBF, funding code [01ZZ1802U]).

References

[1] Kubben P, Dumontier M, Dekker A. Fundamentals of clinical data science. Fundamentals of Clinical Data Science. Springer International Publishing; 2018;1–219.
[2] Nayak A, Božić B, Longo L. Data Quality Assessment of Comma Separated Values Using Linked Data Approach. 2022. p. 240–50.
[3] Overview - FHIR v4.3.0 [Internet]. [cited 2022 Aug 23]. Available from: https://www.hl7.org/fhir/overview.html
[4] Martínez-García A, Cangioli G, Chronaki C, Löbe M, Beyan O, Juehne A, et al. FAIRness for FHIR: Towards Making Health Datasets FAIR Using HL7 FHIR. Stud Health Technol Inform [Internet]. Stud Health Technol Inform; 2022 [cited 2022 Aug 23];290. Available from: https://pubmed.ncbi.nlm.nih.gov/35672963/
[5] Healthcare Integration Engine | Mirth® Connect by NextGen Healthcare [Internet]. [cited 2022 Nov 28]. Available from: https://www.nextgen.com/products-and-services/integration-engine
[6] Deppenwiese N, Delpy P, Lambarki M, Lablans M. ADT2FHIR - A Tool for Converting ADT/GEKID Oncology Data to HL7 FHIR Resources. Stud Health Technol Inform [Internet]. Stud Health Technol Inform; 2021 [cited 2022 Aug 29];283:104–10. Available from: https://pubmed.ncbi.nlm.nih.gov/34545825/
[7] Fette G, Ertl M, Störk S. Translating openEHR Models to FHIR. Stud Health Technol Inform [Internet]. Stud Health Technol Inform; 2020 [cited 2022 Aug 29];270:1415–6. Available from: https://pubmed.ncbi.nlm.nih.gov/32570686/
[8] Pfaff ER, Champion J, Bradford RL, Clark M, Xu H, Fecho K, et al. Fast Healthcare Interoperability Resources (FHIR) as a Meta Model to Integrate Common Data Models: Development of a Tool and Quantitative Validation Study. JMIR Med Inform [Internet]. JMIR Med Inform; 2019 [cited 2022 Aug 29];7. Available from: https://pubmed.ncbi.nlm.nih.gov/31621639/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230062

Implementation Status of the Proposal for a Regulation of the European Health Data Space in Portugal: Are We Ready for It?

Margarida MATEUS* [a] & Maria LOUREIRO* [a,1], A. Raquel FERNANDES [a], Miguel OLIVEIRA [a] and Ricardo CRUZ-CORREIA [a]
[a] *Faculty of Medicine, University of Porto, Porto, Portugal*
*These authors contributed equally to this work.

Abstract. The European Health Data Space (EHDS) proposal aims to establish a set of rules and governance frameworks to promote the use of electronic health data for both primary and secondary purposes. This study aims at analysing the implementation status of the EHDS proposal in Portugal, particularly the points concerning the primary use of health data. The proposal was scanned for the points that gave member states a direct responsibility to implement actions, and a literature review and interviews were conducted to assess the implementation status of these policies in Portugal This study found that Portugal is well advanced in the implementation of policies concerning the rights of natural persons in relation to the primary use of their personal health data, but also identified challenges, which include the lack of a common interoperability framework for the exchange of electronic health data.

Keywords. Data Protection, Electronic Health Data, European Health Data Space, Primary Use of Data, Portugal

1. Introduction

In May 2022, the European Commission (EC) launched the proposal for a Regulation on the European Health Data Space (EHDS) establishing a set of rules, infrastructures, and governance instruments, under a common governance framework, to promote both primary and secondary uses of electronic health data (EHD) [1].

This represents a crucial step in achieving seamless cross-border healthcare in the European Union (EU) and an important opportunity for using the extensive health data which is currently neglected by granting access to researchers and public health authorities to large international electronic health datasets [2]. The adoption of this ambitious project will accelerate the availability and quality of EHD, improve the diagnosis and treatment of patients and contribute to the continuity of care across borders. The EC aims at enforcing the Regulation by 2025 but this deadline depends on its approval and correct implementation, which relies on the engagement of a highly complex multi-stakeholder environment (e.g., member states, health providers, citizens, academia, regulators, as well as industry).

[1] Corresponding Author: Maria Loureiro, E-mail: up201405046@med.up.pt

In this context, it is essential to analyse to what extent member states are ready for the EHDS. It is well known that when it comes to the primary use of EHD, countries have different maturity levels - some have achieved high levels of digitalization and interoperability within their borders, while others are still lacking basic infrastructure. For example, the summaries of patient health records and electronic prescription services exist in two-thirds of EU countries, but only in a few can they be shared across borders.

This paper aims to analyse the implementation status of the EHDS Regulation proposal with regards the points encompassing the primary use of EHD in the Portuguese healthcare context.

2. Methods

The qualitative study was carried out between November and December 2022. First, there was an extensive analysis of the Regulation proposal and a filtering of the article points that gave member states a direct responsibility to implement one or more actions, resulting in the selection of eighteen points on the access to and transmission of personal EHD for primary use, for implementation across five different areas [1].

Secondly, the researchers conducted a detailed assessment of the status of implementation in Portugal of the policies listed in the proposed Regulation. Due to the extensive nature of this text, the project's scope was restricted to the primary use of EHD - patients' rights, access by health professionals and priority categories of data to be shared - to convey more detail to the final assessment, as depicted in Table 1.

This analysis was based primarily on the consultation of Portuguese governmental websites, websites of public and private Portuguese healthcare providers, and national regulatory entities. To complement this information, non-systematic literature consultations were conducted (although there is a scarcity research on this topic), as well as a semi-structured interview with a representative of the Shared Services for the Portuguese Ministry of Health.

Table 1. Adapted from the Regulation's Proposal on the EHDS on primary use of EHD (Chapter II, Section I) [1] - articles and respective points selected for analysis.

Article	Implementation
3 - Rights of natural persons in relation to the primary use of their personal EHD	5. Member States shall: a) establish one or more electronic health data access services at national, regional or local level enabling the exercise of rights referred to in paragraphs 1 and 2; b) establish one or more proxy services enabling a natural person to authorise other natural persons of their choice to access their electronic health data on their behalf.
	7. Member States shall ensure that, when exercising the right to rectification under Article 16 of Regulation (EU) 2016/679, natural persons can easily request rectification online through the electronic health data access services referred to in paragraph 5, point (a), of this Article.
	9. Notwithstanding Article 6(1), point (d), of Regulation (EU) 2016/679, natural persons shall have the right to restrict access of health professionals to all or part of their electronic health data. Member States shall establish the rules and specific safeguards regarding such restriction mechanisms.
4 - Access by health professionals to personal EHD	3. Member States shall ensure that access to at least the priority categories of electronic health data referred to in Article 5 is made available to health professionals through health professional access services. Health professionals who are in possession of recognised electronic identification means shall have the right to use those health professional access services, free of charge.

Article	Implementation
	4. Where access to electronic health data has been restricted by the natural person, the healthcare provider or health professionals shall not be informed of the content of the EHD without prior authorisation by the natural person, including where the provider or professional is informed of the existence and nature of the restricted electronic health data. In cases where processing is necessary in order to protect the vital interests of the data subject or of another natural person, the healthcare provider or health professional may get access to the restricted electronic health data. Following such access, the healthcare provider or health professional shall inform the data holder and the natural person concerned or his/her guardians that access to electronic health data had been granted. Member States law may add additional safeguards.
5 - Priority categories of personal EHD for primary use	1. Where data is processed in electronic format, Member States shall implement access to and exchange of personal electronic health data for primary use fully or partially falling under the following categories: (a) patient summaries; (b) electronic prescriptions; (c) electronic dispensations; (d) medical images and image reports; (e) laboratory results; (f) discharge reports.

3. Results

In Portugal, there is a government-funded web portal aiming to gather essential information from each citizen to improve the provision of healthcare, which comprises a Personal Area (for patients' access) [3]. Another government-funded web portal comprises a Professional Area (for healthcare providers), an International Area (for international access to data) and an Institutional Area [4].

In their personal area, users take on an active role in maintaining, promoting and improving the documented data about their health status, consequently, providing more effective and safer care. This service provides free, immediate access (through authentication) to the patient's emergency contacts, medication, allergies, illnesses, documents and health reports (this data entries can be either registered by the patient or the healthcare professional), and this data can be accessed by other healthcare professionals (different from the ones that registered firstly the data and if the patient authorizes its accessibility) as long as they belong to the national health service (Serviço Nacional de Saúde - SNS). Additionally, patients have the right to request access to their health data (either electronically or on paper) to the health units where it is registered.

As custodians of health information, health units allow users to consult their local medical records; however, there is no access to the patient's centralized health records across all units. Patients can request access either in person or by email. If the patient cannot physically go to the health unit to require access, they can assign this right to a third party. This authorisation, signed by the user, should include the user's and the selected representative's complete identification (name, citizen card number and address) as well as the information to be consulted [5]. (Article 3, Point 5)

In the portal mentioned above, there are twelve items under the privacy settings, allowing its user to set who will have access to their personal health information. They can also be notified of who, when and where their health information was accessed. Should the patient not authorise the sharing of information, this is only reflected in the central data storage system - the local health units where data is registered will continue to have access to their medical history (e.g., consultations, prescriptions, etc.). Vaccination information is always available, as it is Public Health information. (Article 3, Point 9; Article 4, Point 4)

When it comes to the request for rectification of personal health data, the Portuguese legislation states that the data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning them, as well as the right to have incomplete personal data completed, including by means of an additional declaration [6]. However, changes to patient clinical records can only be performed by a family doctor and/or the attending physician in the professional area of the EHR platform. In the case of electronic prescriptions (ePrescriptions), it is not possible to change the data [7]. (Article 3, Point 7)

Regarding the "Professional Area" of the portal, it allows health professionals to access patients' medical data. The complete medical records are not made available through the portal, only summary data concerning the patient's appointments, treatments, diagnosis, and prescriptions. This service is available in all SNS primary and secondary care units, although efforts are in place to allow for the intake of data collected by private providers. (Article 4, Point 3)

The purpose of the "International Portal" is to give doctors from other EU countries the possibility to consult the patient's clinical information. This access is only possible with prior authorisation by the user, encompassing the Patient Health summary and the patient's ePrescriptions. The Health Summary details the patient's allergies, chronic medication, vaccinations, medical and nursing diagnoses, procedures, and medical devices. In the case of ePrescription, the data corresponds to the electronic medical prescription data: prescription number, prescribed drugs, prescription location and identification of the doctor [8]. In Portugal, this service is currently available to facilitate the sharing and reception of data with more than ten EU member states.

However, electronic dispensations, medical images and image reports, laboratory results and discharge reports cannot yet be shared. (Article 5, Point 1)

4. Discussion

It is essential to consider that this study is based on a Regulation Proposal yet to be approved, and it will likely be modified before it is turned into a law. However, the researchers believe that the current analysis still stands as valuable evidence concerning the status of the implementation of digital health policies in Portugal and can help in the definition of the Portuguese digital health roadmap of the next decade. This research has shown that Portugal is well advanced in collecting and sharing EHD within its borders.

In terms of cross-border EHD sharing, Portugal's technological development was, as in other countries, accelerated by the COVID-19 crisis. Similarly, it is expected that the implementation of the Regulation will proceed at different paces depending on the existing national infrastructure. If, on one hand, the lack of existing infrastructure may pose a challenge (e.g., the need to promote professionals' and users' eHealth literacy; to implement software solutions at the hospital level), on the other hand, it may also represent a competitive advantage, as it allows countries to implement processes from scratch, benefiting from the experience (and lessons learned) of countries leading the digital transformation. Having adopted Electronic Health Records (EHR) in the 1990s, Portugal will have to adapt to the new processes proposed in the Regulation once approved. This is no easy task since the provision of services must be continued, and the legacy systems should be interoperable with future solutions. The main challenge of this project was the scarcity of literature (academic or otherwise) on the implementation of digital health initiatives in Portugal. Although there are references to the use of European

funds (namely the Recovery and Resilience Plan) for digital health transformation, there is a need for more detailed information on the course of implementation and objectives of the financed projects. Only then will it be possible to assess the impact of these projects on the implementation of the EHDS in Portugal.

5. Conclusion

Due to the dynamic nature of digital health transformation, monitoring the trend in impacts arising from the EHDS will constitute a crucial part of the action in this domain.

To ensure that the selected policies deliver the intended results and to inform future revisions, it is necessary to monitor and evaluate their implementation. This can be achieved primarily through reporting by digital health authorities and health data access bodies.

This research shows that Portugal is ready for the EHDS Regulation's approval, and it is on a promising path towards its implementation, particularly regarding the primary use of EHD. However, the proposal is still being discussed, and its implementation depends on the regulation approved. This means that the proposal may still be substantially altered, in which case some aspects evaluated in this paper may be rendered inapplicable. In the opinion of the researchers, the proposal, as it stands, could be overly ambitious to be enforceable in all member states in a coordinated manner.

From the collected information and considering the limitations of this study, the researchers suggest that future in-depth studies on this topic should be conducted, with access to further information and expert opinions. Similarly, future research should also analyse the points on the proposed Regulation that refer to the secondary use of health data.

References

[1] EUR-Lex Publications Office of the European Union [Internet]. Strasbourg: European Commission, DG SANTE. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space, COM/2022/197 final, Document 52022PC0197 - EN. [updated 2022 May 3; cited 2022 Nov 26]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197.

[2] European Commission website [Internet]. Brussels: De Keersmaecker S, Cassidy D.. Questions and answers - EU Health: European Health Data Space (EHDS). [updated 2022 May 3; cited 2022 Nov 26]. Available from: https://ec.europa.eu/commission/presscorner/detail/en/QANDA 22 2712.

[3] SPMS. SNS24 portal [Internet]. Lisboa: Serviços Partilhados do Ministério da Saúde (SPMS), E.P.E.; 2022 [updated 2019 Jan 22; cited 2022 Nov 27]. Available from: https://www.sns24.com/.

[4] SPMS. RSE-SIGA portal. Lisboa: SPMS, E.P.E.; 2023 [cited 2022 Nov 30]. Available from: https://rse-siga.spms.min-saude.pt/.

[5] ERS. Entidade Reguladora da Saúde (ERS) website [Internet]. Porto: ERS; 2020 [updated 2020 Nov 5; cited 2022 Nov 26]. Available from: https://www.ers.pt/pt/utentes/perguntas-frequentes/faq/acesso-a-informacao-de-saude/.

[6] Lei N.º 59/2019, 2019 Aug 8. Diário da República N.º 151/2019, Série I [cited 2022 Nov 28] (Portugal). Available from: https://data.dre.pt/eli/lei/59/2019/08/08/p/dre/pt/html.

[7] CNPD. Comissão Nacional de Proteção de Dados (CNPD) website [Internet]. Lisboa: CNPD; 2021. Direito de retificação dos dados [cited 2022 Nov 25]. Available from: https://www.cnpd.pt/cidadaos/direitos/direito-de-retificacao-dos-dados/.

[8] SPMS. SNS 24 website [Internet]. Lisboa: SPMS, E.P.E.; 2022. A minha saúde na Europa [updated 2022 Nov 28; cited 2022 Nov 30]. Available from: https://www.sns24.gov.pt/guia/a-minha-saude-na-europa/entidade-que-assegura-o-servico-em-portugal/.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230063

Health Synthetic Data to Enable Health Learning System and Innovation: A Scoping Review

Shu-Feng TSAO[a,1], Kam SHARMA[a]. Hateem NOOR[a], Alan FORSTER [b] and Helen CHEN[a]
[a] *School of Public Health Sciences, University of Waterloo, Canada*
[b] *Ottawa Hospital, Ottawa, Ontario, Canada*

Abstract. With the recent advancement in the field of machine learning, health synthetic data has become a promising technique to address difficulties with time consumption when accessing and using electronic medical records for research and innovations. However, health synthetic data utility and governance have not been extensively studied. A scoping review was conducted to understand the status of evaluations and governance of health synthetic data following the PRISMA guidelines. The results showed that if synthetic health data are generated via proper methods, the risk of privacy leaks has been low and data quality is comparative to real data. However, the generation of health synthetic data has been generated on a case-by-case basis instead of being scaled up. Furthermore, regulations, ethics, and data sharing of health synthetic data have primarily been inexplicit, although common principles for sharing such data do exist.

Keywords. Synthetic data, data governance, data sharing, FAIR, CARE

1. Introduction

Health data, especially electronic medical records (EMRs), are often stored in disparate systems and formats, rendering integration and standardization difficult [1-2]. Additionally, health data has been strictly regulated by laws, including the Health Insurance Portability and Accountability Act (HIPAA) in the United States (US), the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, and General Data Protection Regulations (GDPR) in the European Union (EU) [3]. Researchers and developers often depend on de-identified, aggregated data to test theories, models, algorithms, or prototypes, but it takes a substantial amount of time and resources to retrieve, aggregate, and de-identify relevant data before it can be used [1-2]. One approach to solve this issue is the creation of realistic, high-quality synthetic health datasets that capture as many of the complexities of the original data sets, but do not include any real patient data [1]. In this scoping review, synthetic data is considered different from deidentified, aggregated data. The latter remains to be a type of real data, whereas the former is completely unreal data created from the real data. For example, the Clinical Practice Research Datalink (CPRD) in the United Kingdom (UK) has created synthetic datasets available for research [4]. The Agency for Healthcare Research and

[1] Corresponding Author: Shu-Feng TSAO, E-mail: s7tsao@uwaterloo.ca.

Quality (AHRQ) in the US also has Synthetic Healthcare Database for Research (SyH-DR) available [5]. Synthetic health data can reflect the characteristics of a population of interest and be a useful resource for researchers, health information technology developers, and informaticians. Therefore, health synthetic data provides great promises to protect patient privacy, diversify datasets, and enhance medical and innovative research. Unlike UK and US, Canada has very limited sharable and useful high-quality health synthetic datasets that meet findable, accessible, interoperable, and reusable (FAIR) standards, despite its footprint in the Common Infrastructure for National Cohort in Europe, Canada, and Africa (CINECA) projects [6]. Although there are principles such as FAIR and CARE (Collective benefit, Authority to control, Responsibility, Ethics) [7-8], applications or implementations of these principles on health synthetic data have remained limited.

With the advance of machine learning (ML) and artificial intelligence (AI), generation of synthetic data has been extensively studied [9]. Generative Adversarial Networks (GAN), along with its customizations, have been promising methods for synthetic data generation recently [9]. Although GANs still have its limitations, they can preserve privacy of health synthetic data more than conventional statistical methods [9]. Furthermore, federated learning is another promising technique to protect data privacy and security since it can train AI models without exchanging real or synthetic data across multiple nodes or networks. This can prevent critical data compromises, but it has not been optimized and implemented at a larger scale [10-11]. If GANs and federated learning can be used together to generate FAIR or CARE health synthetic data, it will create a robust and optimal health data network to protect sensitive patient data and accelerate health research and innovations [10-11]. Although scholars have thoroughly investigated methods for synthetic data generation [9], other gaps, including data utility and governance of health synthetic data, have not been studied comprehensively. Therefore, this scoping review aimed to better understand the current knowledge in the identified gaps and future directions for the health synthetic data.

2. Methods

The scoping review was completed by following the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [12]. PubMed, Scopus, and Google Scholar were used to search not only peer-reviewed journal articles, but also grey literature related to our research. The primary reviewer screened the titles and abstracts of all possibly relevant articles written in English and published between 2012 to December 2022 to determine whether they should be included in full article reviews and retrieved the full articles. Relevant articles were then read, and each paragraph was coded for specific themes (e.g., DG for data governance) by the primary and secondary reviewers. The entire article was then classified by the majority of paragraph themes. The articles were therefore grouped based on their main theme to summarize the main findings. Any discrepancies were discussed and solved by all the reviewers.

3. Results

3.1. Evaluations of Data Quality, Privacy, and Utility

Synthetic Data generation models, utility of the generated health data and privacy concerns of that synthesized health data are co-related. Currently there is no industry standard to produce health synthetic data, however, one of the most popular models are the GANs. These models produce robust synthetic data when the real-world data is also robust by identifying trends in the real-world data without overfitting the synthesized data [13]. Overfitting can take place when the data generated is too similar or almost identical to the real-world data. This becomes problematic with privacy preservation as some examples could be synthesized that are too similar to real-world data.

With the recent boom in electronically stored health related data, there has been a proportional increase in concerns about privacy protection [14]. Synthetic health data synthesis is a key factor in elevating stress associated with health data related privacy concerns. Currently, many models exist to synthesize synthetic data; however, GANS are the most popular.

Common use cases of health synthetic data can be assigned into 6 general categories: (1) EMRs [15], (2) health insurance claims [5, 13, 16], (3) Administrative heath data or surveys [13-14, 16-18], (4) bioinformatics [6], (5) medical images [15], and (6) sensor data [19]. Depending on how data processing is done, data in these categories can be treated as longitudinal or cross-sectional in corresponding analyses.

3.2. Health Synthetic Governance, Data Sharing, and Ethics

Compared to deidentified real patient data, health synthetic data have primarily remained as a grey area in corresponding regulations that govern and protect patient privacy, and its generation and sharing have also been done on a case-by-case basis for research. This has raised many legal and ethical questions that have no clear answers yet. Take informed consents for example. Under HIPAA's privacy rule in the US, creating deidentified data is regarded as healthcare operations of a covered entity [20]. Therefore, informed consents from patients are not required even if the deidentified data will function as a database for research [20]. The similar logic applies to EU's GDPR and Canada' PIPEDA [3]. However, health synthetic data is not de-identified data. Instead, they are fake data artificially created if properly generated, but they closely reflect characteristics of real data. Therefore, this brings up a question: should synthetic health data be considered as protected health information (PHI) or human subject, thus needing informed consents and/or research ethics reviews?

Although health synthetic data appear to be promising for health innovations, sharing synthetic data health is not as common as established databases consisting of real data. Some have advocated that health synthetic data should also follow the FAIR principles for data sharing and open science [8, 21]. Additionally, CARE principles have gained attractions when indigenous data are involved [7, 22]. Existing histories regarding the unfair and unethical treatments of indigenous peoples have strained relationships between indigenous peoples and researchers, resulting in policies that limit data sharing [23]. However, this exclusion of indigenous data sets poses a limitation for a field such as synthetic health data as indigenous datasets can inform many of the machine learning models and clinical algorithms used in research and training [23]. The underrepresented

sample of indigenous datasets limits the predictive accuracy of machine learning models, which leads to unintended biases and misinformed data decisions for indigenous peoples and their health [23]. Furthermore, data that has been historically available for indigenous peoples tends to focus on negative outcomes. To increase healthcare access and equity for indigenous peoples, a need for accurate indigenous data is necessary [24]. Synthetic health data will help close this gap in knowledge but requires partnerships with indigenous peoples and synthetic data stakeholders to address this limitation. To inform healthcare and data decisions for indigenous peoples, there is strong need for indigenous data sovereignty. One such way to tackle this issue is with data governance that can be implemented in partnership with indigenous peoples [25]. This also further pushes the discussion of indigenous data sovereignty for synthetic health data. It is unclear who would own that data and in which ways indigenous peoples are involved in the data governance process. Nonetheless, the CARE principles can address historical inequities and provide indigenous peoples a platform wherein they have data sovereignty [7, 22].

4. Discussion

As shown in this scoping review, the existing literature about synthetic health data governance and evaluations is scarce. This has suggested gaps to be filled in the future. To generate high-quality and useful health synthetic data, it is important to avoid "garbage in, garbage out." Therefore, the quality of original real data is of great importance. Once synthetic health data are created via proper methods, the risk of privacy breaches becomes lower than deidentified, aggregated real data. However, synthetic data hasn't been generated routinely as a means to share data.

Compared to advanced ML techniques to generate high-quality synthetic health data on a case-by-case basis, data governance, including regulations, ethics, and data sharing, for synthetic health data have remained scarce. Researchers have recommended to follow and apply existing legal and ethical governance, as well as common principles for synthetic health data sharing. However, policies still need to be updated accordingly to explicitly indicate whether or not synthetic health data will be governed as human subject data.

5. Conclusion

Health synthetic data offers a promising solution to accelerate health research and innovations. However, its generations and uses have not been scaled up. Further research and regulatory guidelines are needed in data governance and quality evaluations.

References

[1] Kokosi T, De Stavola B, Mitra R, Frayling L, Doherty A, Dove I, et al. An overview on synthetic administrative data for research. Int J Popul Data Sci [Internet]. 2022;7(1). doi:10.23889/ijpds.v7i1.1727

[2] Kokosi T, Harron K. Synthetic data in medical research. BMJ Med [Internet]. 2022;1(1):e000167. doi:10.1136/bmjmed-2022-000167

[3] Shapiro J. Why digital privacy is so complicated [Internet]. Progressive Policy Institute. 2022 [cited 2023 Jan 5]. Available from: https://www.progressivepolicy.org/publication/why-digital-privacy-is-so-complicated/

[4] Synthetic data [Internet]. Cprd.com. [cited 2023 Jan 5]. Available from: https://cprd.com/synthetic-data
[5] Synthetic healthcare database for research (SyH-DR) [Internet]. Ahrq.gov. [cited 2023 Jan 5]. Available from: https://www.ahrq.gov/data/innovations/syh-dr.html
[6] CINECA - common infrastructure for national cohorts in Europe, Canada, and Africa [Internet]. CINECA. [cited 2023 Jan 5]. Available from: https://www.cineca-project.eu/
[7] Carroll SR, Herczog E, Hudson M, Russell K, Stall S. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Sci Data [Internet]. 2021;8(1):108. Available from: doi:10.1038/s41597-021-00892-0
[8] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data [Internet]. 2016;3:160018. doi:.1038/sdata.2016.18
[9] Hernandez M, Epelde G, Alberdi A, Cilla R, Rankin D. Synthetic data generation for tabular health records: A systematic review. Neurocomputing [Internet]. 2022;493:28–45. doi:10.1016/j.neucom.2022.04.053
[10] Antunes RS, André da Costa C, Küderle A, Yari IA, Eskofier B. Federated learning for healthcare: Systematic review and architecture proposal. ACM Trans Intell Syst Technol [Internet]. 2022;13(4):1–23. doi:10.1145/3501813
[11] Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated Electronic Health Records. Int J Med Inform [Internet]. 2018;112:59–67. doi:10.1016/j.ijmedinf.2018.01.007
[12] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med [Internet]. 2018 [cited 2023 Jan 5];169(7):467–73. doi:10.7326/M18-0850
[13] Bhanot K, Qi M, Erickson JS, Guyon I, Bennett KP. The problem of fairness in synthetic healthcare data. Entropy (Basel) [Internet]. 2021;23(9):1165. doi:10.3390/e23091165
[14] El Emam K, Mosquera L, Jonker E, Sood H. Evaluating the utility of synthetic COVID-19 case data. JAMIA Open [Internet]. 2021;4(1):ooab012. doi:10.1093/jamiaopen/ooab012
[15] Reiner Benaim A, Almog R, Gorelik Y, Hochberg I, Nassar L, Mashiach T, et al. Analyzing medical research results based on synthetic data and their relation to real data results: Systematic comparison from five observational studies. JMIR Med Inform [Internet]. 2020;8(2):e16492. doi:10.2196/16492
[16] Yale A, Dash S, Bhanot K, Guyon I, Erickson JS, Bennett KP. Synthesizing quality open data assets from private health research studies. In: Business Information Systems Workshops. Cham: Springer International Publishing; 2020. p. 324–35. doi:10.1007/978-3-030-61146-0_26
[17] Johnson AEW, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data [Internet]. 2016;3(1):160035. doi:10.1038/sdata.2016.35
[18] Goncalves A, Ray P, Soper B, Stevens J, Coyle L, Sales AP. Generation and evaluation of synthetic patient data. BMC Med Res Methodol [Internet]. 2020;20(1). doi:10.1186/s12874-020-00977-1
[19] Norgaard S, Saeedi R, Sasani K, Gebremedhin AH. Synthetic sensor data generation for health applications: A supervised deep learning approach. Annu Int Conf IEEE Eng Med Biol Soc [Internet]. 2018;2018:1164–7. doi:10.1109/EMBC.2018.8512470
[20] Nass SJ, Levit LA, Gostin LO, Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule. HIPAA, the Privacy Rule, and its application to health research. Washington, D.C., DC: National Academies Press; 2009. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9573/
[21] Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inf Serv Use [Internet]. 2017 [cited 2023 Jan 6];37(1):49–56. doi: 10.3233/ISU-170824
[22] 1Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, et al. The CARE principles for indigenous data governance. Data Sci J [Internet]. 2020;19. doi:10.5334/dsj-2020-043
[23] Boscarino N, Cartwright RA, Fox K, Tsosie KS. Federated learning and Indigenous genomic data sovereignty. Nat Mach Intell [Internet]. 2022;4(11):909–11. doi:10.1038/s42256-022-00551-y
[24] Walker J, Lovett R, Kukutai T, Jones C, Henry D. Indigenous health data and the path to healing. Lancet [Internet]. 2017;390(10107):2022–3. doi:10.1016/s0140-6736(17)32755-1
[25] Love RP, Hardy B-J, Heffernan C, Heyd A, Cardinal-Grant M, Sparling L, et al. Developing data governance agreements with Indigenous communities in Canada: Toward equitable tuberculosis programming, research, and reconciliation. Health Hum Rights. 2022;24(1):21–33. PMID: 35747272; PMCID: PMC9212824.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230064

Reproducibility in 2023 - An End-to-End Template for Analysis and Manuscript Writing

Jonathan M. MANG[a,1], Hans-Ulrich PROKOSCH[a,b] and Lorenz A. KAPSNER[a,c]
[a] *Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Erlangen, Germany*
[b] *Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany*
[c] *Institute of Radiology, Universitätsklinikum Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany*

Abstract. Reproducibility imposes some special requirements at different stages of each project, including reproducible workflows for the analysis including to follow best practices regarding code style and to make the creation of the manuscript reproducible as well. Available tools therefore include version control systems such as Git and document creation tools such as Quarto or R Markdown. However, a re-usable project template mapping the entire process from performing the data analysis to finally writing the manuscript in a reproducible manner is yet lacking. This work aims to fill this gap by presenting an open source template for conducting reproducible research projects utilizing a containerized framework for both developing and conducting the analysis and summarizing the results in a manuscript. This template can be used instantly without any customization.

Keywords. Reproducibility, Data sharing, Template, Container, Computer and information sciences

1. Introduction

The goal of Data Science is to generate new insights from data and encompasses a wide spectrum of data handling procedures, from the initial data collection further on to data processing and preparation, as well as data exploration and statistical modeling, finally to visualization, interpretation and publishing the results [1]. The breadth and complexity of data science scenarios requires comprehensive expertise and experience in a wide range of related topics and domains, e.g. mathematics, statistics, computer science, software development, and domain knowledge of the problem to solve, besides others. Simultaneously, there is increasing interest in automating parts of the data science process, if not the entire process [1]. Additionally, researchers from a number of disciplines in the computer sciences call for *reproducibility* or *reproducible research* as a minimum achievable standard for assessing the value of scientific claims, especially when full independent replication of a study is not possible, for example, due to lack of resources such as time and money [2–5].

[1] Corresponding Author: Jonathan M. Mang, E-mail: jonathan.mang@uk-erlangen.de

Reproducibility in this context refers to obtaining consistent results using the same input data, calculation steps, methods, and codes, as well as the same analysis conditions [6]. As early as 1992, there were detailed ideas and concepts for realizing the goal of reproducible science in which authors are able to link the underlying data, parameters, and programs to each figure in the manuscript so that recalculations can be initiated with the click of a button [7]. In addition, the constantly increasing availability of (big) data coupled with rising computing power is enabling ever more complex, automated as well as multi-site analyses [8–10]. Although the intention to share data and code openly is widely established and a variety of methods and tools have been published to accomplish reproducibility in all steps of scientific research, an easy-to-use and publicly available project template is yet missing to organize reproducible end-to-end analysis including the final manuscript for publishing the results.

Since reproducible deployments are common in software development and the creation of reproducible reports is common in the statistical area, these concepts should be merged and the runtime environment in which the analysis was performed should also be provided in a reproducible manner to ensure the reproducibility of the entire workflow besides the publication of data and code.

This project template provides such a framework in which both the execution of the analysis and creation of the manuscript remain reproducible, including the entire software environment required and its dependencies using state-of-the-art tools.

2. Methods

Many tutorials, tools, and reports are available to facilitate approaching the topic of reproducibility in science [5,11,12]. These reports present in detail tool collections and concepts that can be used within the different stages of an analysis: A good data management is useful to structure all documents and materials as a basis for effective planning of the project which is addressed with a predefined folder structure in the template. Since not only the documents and files, but also the program code should be neat and traceable, as well as regularly backed up to different locations, GitHub was used for decentralised versioning the program code, files and folders in the presented template.

Intermediate steps of an analysis should be clearly traceable with intermediate results [11]. Therefore, all raw data is available in a usable form (accessible, digital and non-proprietary) within the template. Ideally, the rationale for decisions made during the analysis should be recorded and traceable as well. To this end, a small, annotated demo analysis has been integrated into the template, describing the data flow in the code up to the final calculation results embedded in the manuscript. The code follows a clean, consistent style that is intended to make the code easier to read. Additional repetitive tasks are automated using functions.

Some published methods and guidance on how reproducible research is possible also consider the benefits of containerizing the analysis [11]. Thus, with appropriate encapsulation of the scripts and the development environment, it can be ensured that not only the text, the underlying data and the scripts necessary to reproduce the results are available, but also that the complete computational and system environment, including all system and package dependencies, can be used in a defined, versionable condition to

reproduce the results independently of the local system configuration and time. In the provided template, Docker [13,14] was used to implement containerization.

Even after analysis, the container-based setup can be used to summarize the results in a reproducible manuscript format, eliminating the need for manual copying and pasting of data into text or tables, or manual creation of figures. From data analysis to a dynamic document embedding the results, there are many tools to support the publication workflow. In the following, we use Quarto [15], a modern open source scientific and technical publishing system built on Pandoc. Various languages such as R [16] and Python [17] can be used in conjunction with Markdown [18], to create a publication-ready manuscript while adhering to accepted standards.

3. Results

At first, it is sufficient to have Docker [13] and Git [19] installed locally to use the presented framework. The repository of the template (see section "Declarations") contains the structure for an ordered data storage, the code necessary for the analysis, as well as a prepared and directly usable container image, in which the analysis can be performed directly, independent of the local operating system. The container image provided ensures that all packages and system requirements necessary to perform the analysis and display the results are uniformly bundled and identical for each execution.

Next, when analyzing the data, the containerized development environment *RStudio* [20] provides an interface to various programming languages as well as standard code development tools such as auto-completion, syntax highlighting, a variable browser and plug-ins, as well as publishing tools (see Figure 1).

The analysis code, the manuscript (including the literature file) and the container information are part of the code repository and are therefore version controlled. Collaboration with co-authors is easily managed using common Git-based code management tools such as branches and merge requests.

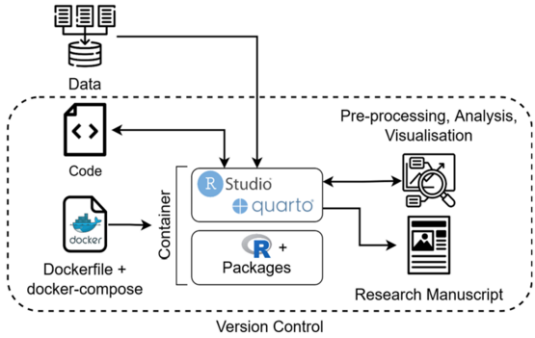


Figure 1. Exemplary workflow to ensure reproducibility in data analysis and manuscript writing. The code for pre-processing, analysis and presentation of results in the manuscript ('code') is version controlled. The build information of the container for the runtime and development environment ('Dockerfile + docker-compose') is also version controlled. The container image is also version controlled and stored in a dedicated online repository. The raw data itself is not version controlled via Git.

Finally, once the analysis is complete, details such as the correct formatting of sources and rendering into the correct output format of the target journal need to be done. For this purpose, Quarto offers a way to integrate journal templates [21] or journal-

specific citation styles [22] without having to change the manuscript text. Once all the details used for the analysis and evaluation as well as the Dockerfile required to create the container with the underlying runtime environment have been stored in the repository, they are shared publicly via the version control system used. To enable reproducibility, the container image with all dependencies is stored in a publicly available package repository. Further details and instructions on all the necessary steps are described in the template.

4. Discussion

We here present a project template that includes currently available tools that are required for conducting reproducible research. Besides the version control of the analysis code and the writing of the manuscript, reproducibility is here extended also to the runtime environment in which the former are performed. The presented template uses GitHub [23] for version control, R [16] as the main programming language, RStudio [20] as the corresponding development environment, and Docker [13] for containerization. Nevertheless, further technologies may be used instead.

Completely reproducible analyses requires all data to be available so that the same results can then be obtained with the analysis steps, which need to be published as well. However, often the raw data cannot be published due to various reasons (storage requirements, data protection, trade secrets). In this case, aggregated data or synthetic data with a structure identical to the original data could be provided instead.

One way to protect intellectual property may be to place restrictions on the permissible use of the dataset in a license or as part of a formal agreement between the researcher and an applicant [3].

5. Conclusions

The purpose of this article is to summarize current state-of-the-art technologies to ensure reproducibility in research and to provide a ready-to-use project template for this purpose. In addition to the methods described for publishing the data and code, this template provides a framework for extending the concepts of reproducibility to the summary of the final results in the form of a manuscript. For this purpose, both a Quarto template that connects the manuscript text with the analysis code and dynamically embeds the results as well as the underlying runtime environment in the form of a Docker container are provided. Together with the publicly available open source software framework used for this purpose, this enables timeless reproducibility of research results.

Declarations

Availability of data and materials: The template repository repub is available on GitHub (https://github.com/joundso/repub). Extensive information about the usage and optional parameterization of the template is available in the respective readme files in the repository.

Funding: This work was partly funded by the German Federal Ministry of Education and Research (BMBF) within the Medical Informatics Initiative (MIRACUM Consortium) under the Funding Number FKZ: 01ZZ1801A.

Authors' contributions: JMM and LAK set up the template. JMM wrote the original manuscript. LAK and HUP supervised the project, reviewed, and edited the manuscript. All authors read and approved the final version. The present work was performed in (partial) fulfillment of the requirements for obtaining the degree "Dr. rer. biol. hum." at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (JMM).

Competing interests: None declared.

References

[1] De Bie T, De Raedt L, Hernández-Orallo J, Hoos HH, Smyth P, Williams CK. Automating data science. Communications of the ACM. 2022 Feb 23;65(3):76-87.

[2] Schwab M, Karrenbach N, Claerbout J. Making scientific computations reproducible. Computing in Science & Engineering. 2000 Nov;2(6):61-7.

[3] Laine C, Goodman SN, Griswold ME, Sox HC. Reproducible research: moving toward research the public can really trust. Annals of Internal Medicine. 2007 Mar 20;146(6):450-3.

[4] Gentleman R. Reproducible Research: A Bioinformatics Case Study, Statistical Applications in Genetics and Molecular Biology. 4 (2005). doi:10.2202/1544-6115.1034.

[5] Peng RD. Reproducible Research in Computational Science, Science. 334 (2011) 1226–1227. doi:10.1126/science.1213847.

[6] National Academies of Sciences, and Medicine, Reproducibility and Replicability in Science, The National Academies Press, Washington, DC, 2019. doi:10.17226/25303.

[7] Claerbout JF, Karrenbach M. Electronic documents give reproducible research a new meaning. InSEG technical program expanded abstracts 1992 1992 Jan 10 (pp. 601-604). Society of Exploration Geophysicists.

[8] Prokosch HU, Bahls T, Bialke M, Eils J, Fegeler C, Gruendner J, Haarbrandt B, Hampf C, Hoffmann W, Hund H, Kampf M. The COVID-19 data exchange platform of the German university medicine. InChallenges of Trustable AI and Added-Value on Health 2022 (pp. 674-678). IOS Press.

[9] Kapsner LA, Kampf MO, Seuchter SA, Gruendner J, Gulden C, Mate S, Mang JM, Schüttler C, Deppenwiese N, Krause L, Zöller D. Reduced rate of inpatient hospital admissions in 18 German university hospitals during the COVID-19 lockdown. Frontiers in public health. 2021:1018.

[10] Schüttler J, Mang JM, Kapsner LA, Seuchter SA, Binder H, Zöller D, Kohlbacher O, Boeker M, Zacharowski K, Rohde G, Balig J. Letalität von Patienten mit COVID-19: Untersuchungen zu Ursachen und Dynamik an deutschen Universitätsklinika. Anästhesiologie & intensivmedizin. 2021;62:244-57.

[11] Alston JM, Rick JA. A beginner's guide to conducting reproducible research. Bulletin of the Ecological Society of America. 2021 Apr 1;102(2):1-4.

[12] Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS computational biology. 2013 Oct 24;9(10):e1003285.

[13] Merkel D. Docker: lightweight linux containers for consistent development and deployment. Linux j. 2014 Mar 2;239(2):2.

[14] Bernstein D. Containers and cloud: From lxc to docker to kubernetes. IEEE cloud computing. 2014 Sep;1(3):81-4.doi:10.1109/MCC.2014.51.

[15] Quarto, RStudio, PBC, 2022. https://quarto.org/ (accessed December 5, 2022).

[16] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2022. https://www.R-project.org/.

[17] G. Van Rossum, and F.L. Drake Jr, Python reference manual, Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[18] Daring Fireball: Markdown Syntax Documentation, (n.d.). https://daringfireball.net/projects/markdown/syntax#philosophy (accessed December 27, 2022).

[19] S. Chacon, and B. Straub, Pro git, Apress, 2014.

[20] Posit team, RStudio: Integrated development environment for R, Posit Software, PBC, Boston, MA, 2022. http://www.posit.co/.

[21] Quarto Journal Templates, (2023). https://github.com/quarto-journals (accessed January 9, 2023).

[22] Citation Style Language, (2023). https://citationstyles.org/ (accessed January 9, 2023).

[23] github, GitHub, (2020). https://github.com/.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230065

Assessing the FAIRness of Deep Learning Models in Cardiovascular Disease Using Computed Tomography Images: Data and Code Perspective

Kirubel Biruk SHIFERAW[a,1], Atinkut ZELEKE[a] and Dagmar WALTEMATH[a]
[a] *Medical Informatics Laboratory, Institute for Community Medicine, University Medicine Greifswald, Germany*

ORCiD ID: Kirubel Biruk Shiferaw https://orcid.org/0000-0002-7411-1411
Atinkut Zeleke https://orcid.org/0000-0001-7838-9050
Dagmar Waltemath https://orcid.org/0000-0002-5886-5563

Abstract. The interest in the application of AI in medicine has intensely increased over the past decade with most of the changes in the past five years. Most recently, the application of deep learning algorithms in prediction and classification of cardiovascular diseases (CVD) using computed tomography (CT) images showed promising results. The notable and exciting advancement in this area of study is, however, associated with different challenges related to the findability (F), accessibility(A), interoperability(I), reusability(R) of both data and source code. The aim of this work is to identify reoccurring missing FAIR-related features and to assess the level of FAIRness of data and models used to predict/diagnose cardiovascular diseases from CT images. We evaluated the FAIRness of data and models in published studies using the RDA (Research Data Alliance) FAIR Data maturity model and FAIRshake toolkit. The finding showed that although AI is anticipated to bring ground breaking solutions for complex medical problems, the findability, accessibility, interoperability and reusability of data/metadata/code is still a prominent challenge.

Keywords. FAIR Principles, Deep learning, cardiovascular disease, computed tomography, RDA FAIR Data maturity model

1. Introduction

The development of computational models using Artificial intelligence (AI) in Medicine has gained high interest in the last five years due to the new possibilities to incorporate multi-modal biomedical data as well as to mimic and to explore the complexity of the events and interdependencies at various levels (molecular, cellular, tissue/ organ, whole-body) of the human biomedical systems [1]. This development has opened new paths in approaching medical problems with complex and robust AI

[1] Corresponding Author: Kirubel Biruk Shiferaw, Walther-Rathenau-Str. 48, Medical Informatics Laboratory, University Medicine Greifswald, D-17475 Germany; E-mail: s-kishif@uni-greifswald.de.

applications in terms of virtual (model) and physical (device) methods. Deep learning (DL) revolutionized the application of AI in medicine, especially in image processing [2].

Most recently, the application of DL algorithms in cardiovascular disease (CVD) risk/event predication and classification using SPECT/CT (A single-photon Emission Computed Tomography) and PET/CT (Positron Emission Tomography) images showed promising results [3]. Given the advancement in scanner technology in both image quality and dimensions, Computed Tomography (CT) is well suited for advanced image analysis using deep neural networks [4]. Indeed, the published research outputs show an increase of approaches for predicting and diagnosing CVD using CT imaging.

However, the reproducibility of DL-based studies has become a challenge [5]. It is consistently mentioned in the literature that reproducibility is a core issue in the scientific process, and this includes AI research [6]. The notable and exciting advancement in this area of study is associated with different challenges relating to the findability (F), accessibility(A), interoperability(I), reusability(R) of both data and source code.

The FAIR guiding principles are one of the recently applied set of guidelines to facilitate the discovery and reuse of scientific digital objects including data, metadata, software and tools [7]. Substantial effort by different working initiatives has been made to quantify FAIRness of digital objects [8], and the FAIR evaluation tools have been developed to help identify weak points in data and code representation within particular scientific domains (e.g. https://fairassist.org/#!/) [9]. Beside contributing to the reproducibility of published studies, the sharing of data and code facilitates rigor scientific practice and reassures the validity of the claimed results [10]. Using standard frameworks, adhering to guiding principles, and developing and reporting guidelines are some of the most common approaches of standardizing data and code sharing, e.g. as demonstrated by the "Computational Modeling in Biology" Network (COMBINE: https://co.mbine.org/) community in the area of computational biology [11, 12].

We aim to assess the level of FAIRness of current DL models in imaging in CVD, and to identify systemic lacks of the FAIR principles which might lead us to recommend coordinated actions for future research in the field. Therefore, we used the RDA (Research Data Alliance) FAIR Data maturity model [13] and FAIRshake tools [14] to evaluate the FAIRness of DL models and associated data in studies on CVDs from CT images.

2. Method

First, we defined the following set of keywords to describe the CVD-applied DL models of interest: Cardiovascular, CT scan, Deep learning, Diagnosis, Heart Defect, Computed, tomography, Hierarchical learning, classification, Congenital, X-ray computed, prognosis and prognosis. After keywords had been identified, using this keyword list, we performed a comprehensive search in Web of Science within the time frame of 2016-2022. We included studies conducted with DL models to predict/classify cardiovascular disease/event from PET(CT)/SPCT(CT) images. As such, a list of 109 publications with their corresponding models was used for further descriptive analysis and FAIRness assessment using essential indicators of the RDA FAIR Data maturity model and the FAIRshake tool. The RDA FAIR Data maturity model is an evaluation tool to assess adherence to the FAIR principles. The available indicators provide three different

degrees with respect to impact on the FAIRness, namely, Essential (critical to achieve FAIRness), Important (with high contribution to the FAIRness features), or Useful (increase the overall FAIRness level of the resources) [15]. The RDA FAIR Data model furthermore offers a scale-based approach to prioritize and self-evaluate the level of FAIRness. FAIRshake is also another tool developed to facilitate the establishment of community driven metrics and rubrics paired with manual and automated FAIR assessment with insignia visualization [14].

3. Results

The search resulted in 109 articles. After excluding studies with closed access, in non-English language, and studies that were irrelevant for our objective, only 22 studies were further analyzed. With respect to datasets associated with the respective studies, only 5/22 (22.7%) provide a URL for the data; 7/22 (31.8%) provide neither metadata nor URL at all in the document. Furthermore, only 4/22 (18.2%) studies provide information on how to access the data used in the study such as "Available on reasonable request from the Author". Finally, 4/22 (18.2%) studies provided both URL and metadata. With respect to code availability, only 2/22 (9.1%) studies made their code available for reuse with license details and only one study stated that the code is "available on reasonable request". Only 2/22 (9.1%) of the studies used a reporting standard namely TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) [16] and STARD (The Standards for Reporting of Diagnostic Accuracy) [17].

The analysis from the RDA-FAIR Data maturity model (Figure 1) shows that most of the essential indicators were not satisfied, particularly metadata related indicators were poorly represented.

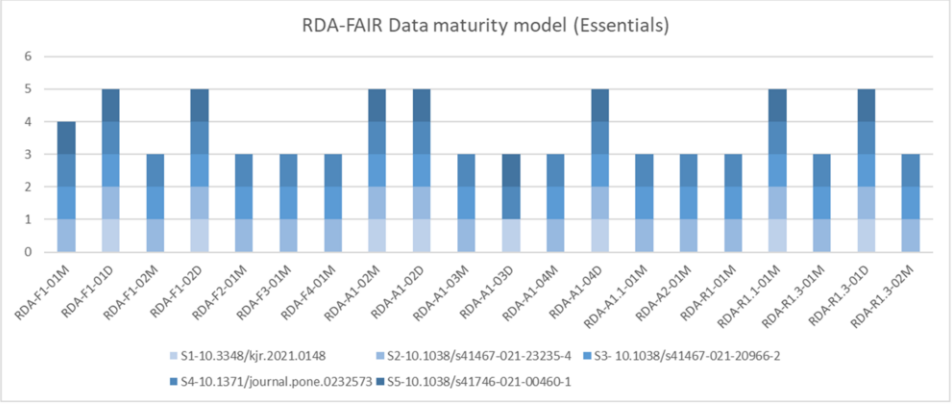


Figure 1. RDA-FAIR data maturity model assessment (x-axis: Essential indicators, y-axis: number of studies satisfying specific indicators)

The FAIRshake insignia visualization (Figure 2) is based on FAIR metrics by FAIRmetrics.org. It also shows that a satisfactory level of FAIRness is still not reached in the scientific field of DL studies on CVD using CT image. Most squares in Figure 2 are red (hence do not satisfy the FAIR metrics) with most of the challenges in Findability and Reusability sectors.

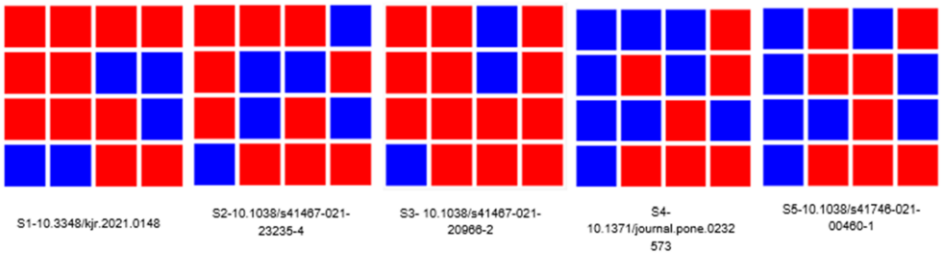


Figure 2. FAIRshake insignia assessment. Each set of four squares of the insignia from the left upper corner to right lower corner represents one aspect of FAIR [(F, A, I&R)].

4. Discussion

The advancement in AI applications, especially DL algorithms in the medical and healthcare domains, is becoming a promising asset for precise and personalized care [18]. However, our first assessment of the *status quo* shows that the studies evaluated on this work do not share their data or code/software; the studies also do not provide protocols. It would be worth conducting further research to determine the extent to which the lack of code and data sharing is a general trend, rather than specific to the domain of CVD. It is understandable that patient data is sensitive and the developed algorithms are intellectual properties [19]. However, sharing does not necessarily mean providing unlimited access for free, rather using a set of protocols and appropriate licenses that enable other researchers to use and cite the work as needed.

We noticed that some researchers tend to share a URL for their data and code which is a good starting point. However, a dataset/code without detailed metadata is not easily reusable as context information is missing. We argue that researchers should publish a detailed standardized metadata along with their research outcomes to facilitate FAIRness of their resources and to potentially increase the reproducibility [20] and reusability of DL models in CVD. It is also important to note that if results and models are irreproducible, it is very likely that efforts will need to be duplicated, resulting in extra monetary costs, longer time to publication and less trust [21, 22].

It is important to know that a FAIR assessment does not assess the quality of methodology or result of the studies, but it indicates how well a study adheres to modern research data management practices in terms of findability, accessibility, interoperability and reusability of the associated data/code. While a detailed comparison of the functionality of the two methods is beyond the scope of this work, the similar results from both approaches suggest that further research on this topic could be insightful

We therefore recommend biomedical scientists to use available FAIR assessment tools, including the RDA FAIR data maturity model, to evaluate their own works, preferably before publication. Help in using these tools and deriving a data management strategy can be obtained from data stewards at research institutions, data integration centers, or community work groups.

5. Conclusion

Although AI is anticipated to bring ground breaking solutions for complex medical problems, adherence to the FAIR principles for data/metadata/code is still a prominent challenge. Authors should consider standardized ways of sharing their data/metadata/code. Other stakeholders in the publishing ecosystem such as reviewers, editors and publishers should encourage FAIR sharing for the interest of reproducibility, trust and ultimately for the advancement of open science.

References

[1] Shiferaw, K.B., D. Waltemath, and A. Zeleke, Disparities in Regional Publication Trends on the Topic of Artificial Intelligence in Biomedical Science Over the Last Five Years: A Bibliometric Analysis. Studies in Health Technology and Informatics, 2022. **294**: p. 609-613.
[2] Fu, Y., et al., A review of deep learning based methods for medical image multi-organ segmentation. Physica Medica, 2021. **85**: p. 107-122.
[3] Slart, R.H., et al., Position paper of the EACVI and EANM on artificial intelligence applications in multimodality cardiovascular imaging using SPECT/CT, PET/CT, and cardiac CT. European journal of nuclear medicine and molecular imaging, 2021. **48**(5): p. 1399-1413.
[4] Lin, A., et al., Artificial intelligence in cardiovascular CT: Current status and future implications. Journal of cardiovascular computed tomography, 2021. **15**(6): p. 462-469.
[5] Hartley, M. and T.S. Olsson, dtoolai: Reproducibility for deep learning. Patterns, 2020. **1**(5): p. 100073.
[6] Gundersen, O.E. and S. Kjensmo. State of the art: Reproducibility in artificial intelligence. in Proceedings of the AAAI Conference on Artificial Intelligence. 2018.
[7] Wilkinson, M.D., et al., The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 2016. **3**(1): p. 1-9.
[8] Wilkinson, M.D., et al., Evaluating FAIR maturity through a scalable, automated, community-governed framework. Scientific data, 2019. **6**(1): p. 1-12.
[9] Sansone, S.-A., et al., FAIRsharing as a community approach to standards, repositories and policies. Nature Biotechnology, 2019. **37**(4): p. 358-367.
[10] Peng, R.D., F. Dominici, and S.L. Zeger, Reproducible epidemiologic research. American journal of epidemiology, 2006. **163**(9): p. 783-789.
[11] Waltemath, D., et al., The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE). Journal of integrative bioinformatics, 2020. **17**(2-3).
[12] Niarakis, A., et al., Addressing barriers in comprehensiveness, accessibility, reusability, interoperability and reproducibility of computational models in systems biology. Briefings in Bioinformatics, 2022.
[13] Alliance, R.D. RDA FAIR Data maturity model. [cited 2022; Available from: https://www.rd-alliance.org/groups/fair-data-maturity-model-wg.
[14] Clarke, D.J., et al., FAIRshake: toolkit to evaluate the FAIRness of research digital resources. Cell systems, 2019. **9**(5): p. 417-421.
[15] FAIR Data Maturity Model Working Group, FAIR Data Maturity Model. Specification and Guidelines (1.0), 2020. https://doi.org/10.15497/rda00050.
[16] Collins, G.S., et al., Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of British Surgery, 2015. **102**(3): p. 148-158.
[17] Bossuyt, P.M., et al., Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Annals of internal medicine, 2003. **138**(1): p. 40-44.
[18] Razzak, M.I., S. Naz, and A. Zaib, Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps, 2018: p. 323-350.
[19] Jean-Paul, S., et al. Issues in the Reproducibility of Deep Learning Results. in 2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). 2019.
[20] Hartley, M. and T.S.G. Olsson, dtoolAI: Reproducibility for Deep Learning. Patterns, 2020. **1**(5): p. 100073.
[21] Gundersen, O.E., The Reproducibility Crisis Is Real. AI Mag., 2020. **41**(3): p. 103-106.
[22] König, M., et al. Challenges and opportunities for system biology standards and tools in medical research. in 7th Workshop on Ontologies and Data in Life Sciences, ODLS 2016. 2016. CEUR-WS.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230066

No Transfer Without Validation: A Data Sharing Framework Use Case

Hauke HUND[a,1], Reto WETTSTEIN[b], Christopher HAMPF[c], Martin BIALKE[c], Maximilian KURSCHEIDT[a], Simon T SCHWEIZER[a], Christoph ZILSKE[a], Simon MÖDINGER[a] and Christian FEGELER[a]
[a] *GECKO Institute, Heilbronn University of Applied Sciences, Heilbronn, Germany*
[b] *Institute for Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany*
[c] *Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany*

Abstract. Availability and accessibility are important preconditions for using real-world patient data across organizations. To facilitate and enable the analysis of data collected at a large number of independent healthcare providers, syntactic- and semantic uniformity need to be achieved and verified. With this paper, we present a data transfer process implemented using the Data Sharing Framework to ensure only valid and pseudonymized data is transferred to a central research repository and feedback on success or failure is provided. Our implementation is used within the CODEX project of the German Network University Medicine to validate COVID-19 datasets at patient enrolling organizations and securely transfer them as FHIR resources to a central repository.

Keywords. Validation, Data Sharing, Framework, Open Source, FHIR, BPMN

1. Introduction

To facilitate medical research at German university hospitals, the Network University Medicine (NUM) was created at the onset of the COVID-19 pandemic. One of the projects funded by the Federal Ministry of Education and Research (BMBF) within NUM is the COVID-19 Data Exchange Platform (CODEX) with the goal of harmonizing, collecting, distributing and analyzing real-world patient data.

While the usage of real-world data promises new insides for clinical research, it also brings its own challenges and limitations [1,2], with availability and accessibility of the data being important preconditions. Within the German Medical Informatics Initiative (MII)[2], infrastructure components at university hospitals have been established to provide access to data for researchers, including harmonized data definitions, accessible ontologies and data integration centers, including use and access committees.

The conceptual approach of the CODEX project, building on the infrastructure of the MII, was described by Prokosch, et al. in [3]. With this paper we want to report on

[1] Corresponding Author: Hauke Hund, GECKO Institute, Heilbronn University of Applied Sciences, Max-Planck-Straße 39, 74081 Heilbronn, Germany; E-mail: hauke.hund@hs-heilbronn.de.

[2] https://www.medizininformatik-initiative.de/en/start (accessed 2023 March 4).

the data transfer process deployed, with a focus on data validation as well as feedback on success or failure of data transfers.

2. Methods

The architecture and data transfer process of the CODEX project were developed based on existing tools from the MII as well as requirements defined by the data protection guideline created within the CODEX project.

A process plugin for the Data Sharing Framework (DSF) was implemented to enable automatic patient data validation and transfer. The DSF allows distributed business processes to be executed across organizations, with processes modeled using BPMN 2.0 and data exchange using HL7 FHIR R4. The DSF consists of a FHIR server accessible from other organizations and a private Business Process Engine (BPE) to integrate local and remote systems [4].

The CODEX project utilizes the German Corona Consensus Dataset (GECCO) designed by Sass, et al. [5] with data transfers to a central research repository after informed consent [6].

Components from the HAPI FHIR library[3] were used to implement a client for the federated Trusted Third Party service at Greifswald University Hospital [7], the FHIR terminology server at Köln University Hospital [8], the central research repository's FHIR Bridge [9] and to generate *StructureDefinition* snapshots and validate FHIR resources at patient enrolling organizations.

3. Results

3.1. Requirements and Architecture

Several requirements were defined for the CODEX project's data transfer architecture and process: GECCO data (MDAT) should be stored in FHIR servers only accessible from local networks. The transfer process needs to be able to send complete datasets or only modified resources. Only valid datasets should be transported, with resources considered valid, if they follow the FHIR implementation guide and terminologies defined by the GECCO dataset. Directly identifying information (IDAT) needs to be removed from FHIR resources before transport to the central repository.

Privacy-preserving record linkage based on one-way hashed IDAT must be performed across organizations with unique pseudonyms (PSN_s) for all enrolling organizations and the central repository (PSN_t). The enrolling organizations (Source) need to be hidden from the central repository (Target) to reduce the risk of patient re-identification. The datasets (MDAT) and error messages (Error) need to be encrypted during transport so that they can only be read by the central repository and the enrolling organization.

Figure 1 shows a generalized version of the data transfer architecture used within the CODEX project. Patient enrolling organizations are depicted as *Source* and the central repository as *Target*. Record linkage and generation of organization-specific pseudonyms are performed by the *federated Trusted Third Party* (fTTP), with the *Data*

[3] https://github.com/hapifhir/hapi-fhir (accessed 2023 March 4).

Transfer Hub (DTH) acting as a middle man to hide the sending organization from the receiving central repository and to enforce organization-specific pseudonyms.

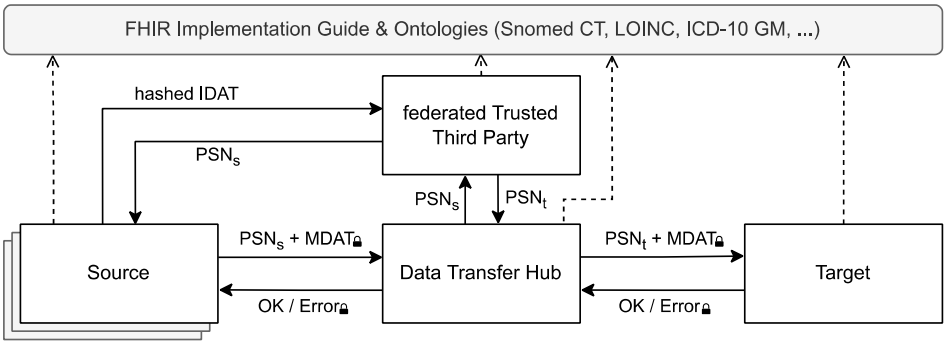


Figure 1. Generalized data sharing architecture.

All connections are based on the FHIR API with transport encrypted using TLS 1.3. Datasets are encrypted asymmetrically at the *Source*- using a public-key (p) from the *Target* organization to hide the content from the *DTH*. A hybrid cryptosystem (RSA+ AES) is used with two AES keys (a, b) generated at the *Source* organization for every data transfer, resulting in the data format $rsa_p(a) + aes_a(b + MDAT)$ for transferring encrypted MDAT and $aes_b(Error)$ for the return of encrypted validation errors.

3.2. Data Transfer Process

At the *Source* organizations datasets are read, pseudonymized, directly identifying information removed, validated and encrypted. Encrypted datasets and pseudonyms (PSN_s) are send to the *DTH*. The *DTH* replaces the pseudonym (PSN_s) with a pseudonym for the *Target* organization (PSN_t) using the *fTTP*. At the *Target* organization, datasets are decrypted, validated and stored. Validation errors that may contain patient information are encrypted for the return path or an Ok message is send back.

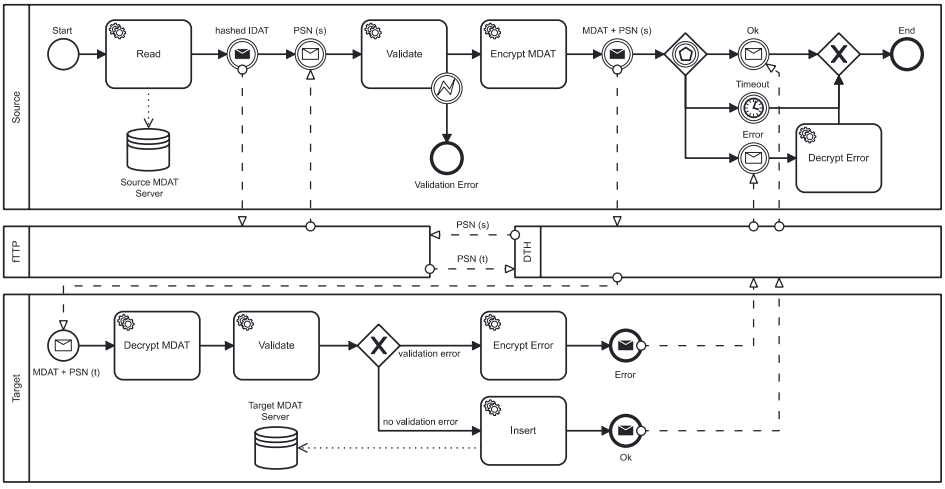


Figure 2. Data sharing process, simplified to improve readability.

Other technical errors at the *fTTP*, *DTH* or *Target* organizations that do not contain patient data are transported back without additional encryption, including high-level codes that enable monitoring. A simplified version of the data transfer process is depicted in Figure 2. The complete processes for the *Source*, *DTH* and *Target* organizations are implemented as a process plugin for the DSF, including executable BPMN models. The process plugin is available as open source under the Apache 2.0 license on GitHub[4].

3.3. Data Validation

Datasets (MDAT) are transferred as FHIR *Bundle* resources of type *transaction*, with individual resources included using *conditional updates*. Before the *Bundle* is encrypted for transport at the *Source* organization, the content is validated against the GECCO FHIR implementation guide[5], with the process failing if validation errors occur.

Preparatory steps need to be performed during the startup of the DSF BPE to complete the resource validation offline without sending patient data to an external terminology- or validation server: First, the implementation guide package and dependent packages are downloaded[6]. Second, required metadata resources are extracted from the downloaded packages. Third, all *ValueSet* resources are *expanded* locally or, if necessary, using the central terminology server. Fourth, profile *snapshots* are calculated for all required *StructureDefinition* resources. And finally, all downloaded and generated resources are stored locally in a file system cache to improve subsequent BPE startups.

While inserting the transported FHIR resources into the central repository, a second validation against the implementation guide is executed, and validation steps across existing FHIR resources of the entire patient collective can be performed.

4. Discussion

Availability and accessibility are important preconditions for clinical research using real-world data. As part of the CODEX project, a data transfer and access platform was created with patients being enrolled at 34 university hospitals in Germany. To allow for the analysis of data collected across various new and existing legacy systems, a common data model was created and enforced by validation.

By validating data at the enrolling organizations and the central repository, we can provide fast feedback to data providers, minimize the number of invalid datasets being transported, and perform validation steps across the entire patient collective.

Although FHIR profiles provide a good mechanism for defining and validating data semantics, there is currently no simple mechanism for specifying resource properties, such as the patient's name or address, which may exist locally but should not be transferred to the central repository. Employing two different but related profiles, one allowing additional properties locally and another strictly enforcing data protection rules, could help but would require the maintenance of two sets of FHIR profiles.

With our process plugin for the DSF data for newly enrolled patients or updates for existing patients can be validated and transferred fully automatically. With the included return channel across the distributed system, we can inform data providers about errors

[4] https://github.com/num-codex/codex-processes-ap1 (accessed 2023 March 4).

[5] https://simplifier.net/forschungsnetzcovid-19 (German, accessed 2023 March 4).

[6] FHIR implementation guide packages are downloaded from https://packages.simplifier.net

during transport or data processing. Because the DSF can send emails when errors occur, operators do not need to monitor the system in production manually.

Using a process plugin for the DSF to validate and transport datasets allowed us to reuse existing infrastructure at German university hospitals and should minimize project-specific maintenance requirements in the future.

5. Conclusions

The data sharing architecture and processes implemented within the German Network University Medicine CODEX project enable the secure transport of real-world patient data using FHIR resources to conduct prospective studies with centralized data storage.

To analyze real-world patient data collected from a large number of independent healthcare providers, syntactic- and semantic uniformity need to be achieved and verified. To improve feedback to data collectors and to minimize the amount of invalid data being transported, data validation should be conducted at the centralized data storage site and also at every patient enrolling organization.

The described implementation enables validation of FHIR resources based on FHIR implementation guides and improves validation speeds while minimizing data transfers.

Acknowledgements

The project is funded by the German Federal Ministry of Education and Research (BMBF, grant ids: 01ZZ1802E, 01ZZ1802A, 01ZZ1801M and 01KX2021).

References

[1] Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data. Onkologe (Berl). 2018;24(Suppl 2):91-98. doi: 10.1007/s00761-018-0358-3

[2] Alemayehu D, Ali R, Alvir JMJ, Cappelleri JC, et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using “Real-World Data”. Journal of Managed Care Pharmacy. 2011;17(9 Supp A):1–37. doi: h10.18553/jmcp.2011.17.s9-a.1

[3] Prokosch H-U, Bahls T, Bialke M, Eils J, et al. The COVID-19 Data Exchange Platform of the German University Medicine. Stud Health Technol Inform. 2022 May;294:674-678. doi: 10.3233/SHTI220554

[4] Hund H, Wettstein R, Heidt CM, Fegeler C. Executing Distributed Healthcare and Research Processes – The HiGHmed Data Sharing Framework. Stud Health Technol Inform. 2021 May;278:126-133. doi: 10.3233/SHTI210060

[5] Sass J, Bartschke A, Lehne M, Essenwanger A, et al. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. BMC Medical Informatics and Decision Making. 2020;20(1). doi: 10.1186/s12911-020-01374-w

[6] MII Consent Working Group. Patient Consent Form Template. 2020 November [cited 2023 March 4] Available from: https://www.medizininformatik-initiative.de/sites/default/files/2020-11/MII_WG-Consent_Patient-Consent-Form_v1.6d_engl-version.pdf

[7] Bahls T, Hampf C, Bialke M, Hoffmann W. Lösungsbaustein fTTP (federated Trusted Third Party) als ein Enabler für vernetzte medizinische Forschung mit dezentraler Datenhaltung. 2022 November [cited 2023 March 4] Available from: https://www.ths-greifswald.de/forscher/num/fttp-fact-sheet

[8] Mateen A, Tielsch-Nebel R, Okereke J. Terminology Server Services Concept. 2021 January [cited 2023 March 4] Available from: https://webstatic.uk-koeln.de/im/dwn/pboxx-pixelboxx-238948/terminology-server-services-concept-uniklinik-koeln.pdf

[9] EHRbase: FHIR Bridge. [cited 2023 March 4] Available from: https://github.com/ehrbase/fhir-bridge

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230067

Comparative Analysis of Electrodermal Activity Decomposition Methods in Emotion Detection Using Machine Learning

Sriram Kumar P[a,1], Praveen Kumar GOVARTHAN [a], Nagarajan GANAPATHY [b] and Jac Fredo AGASTINOSE RONICKOM [a]
[a] *School of Biomedical Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India-221005*
[b] *Department of Biomedical Engineering, Indian Institute of Technology Hyderabad, Kandi, Telangana, India*
ORCiD ID: Sriram Kumar P https://orcid.org/https://orcid.org/0000-0002-8604-2329

Abstract. Electrodermal activity (EDA) reflects sympathetic nervous system activity through sweating-related changes in skin conductance. Decomposition analysis is used to deconvolve the EDA into slow and fast varying tonic and phasic activity, respectively. In this study, we used machine learning models to compare the performance of two EDA decomposition algorithms to detect emotions such as amusing, boring, relaxing, and scary. The EDA data considered in this study were obtained from the publicly available Continuously Annotated Signals of Emotion (CASE) dataset. Initially, we pre-processed and deconvolved the EDA data into tonic and phasic components using decomposition methods such as cvxEDA and BayesianEDA. Further, 12 time-domain features were extracted from the phasic component of EDA data. Finally, we applied machine learning algorithms such as logistic regression (LR) and support vector machine (SVM), to evaluate the performance of the decomposition method. Our results imply that the BayesianEDA decomposition method outperforms the cvxEDA. The mean of the first derivative feature discriminated all the considered emotional pairs with high statistical significance ($p<0.05$). SVM was able to detect emotions better than the LR classifier. We achieved a 10-fold average classification accuracy, sensitivity, specificity, precision, and f1-score of 88.2%, 76.25%, 92.08%, 76.16%, and 76.15% respectively, using BayesianEDA and SVM classifiers. The proposed framework can be utilized to detect emotional states for the early diagnosis of psychological conditions.

Keywords. Emotion detection, Electrodermal activity, Deconvolution, Time-domain features, Machine learning.

1. Introduction

Electrodermal activity (EDA) is a physiological measure of changes in the sympathetic system, reflecting emotional and cognitive states. It tracks the changing electrical conductance of the skin due to the activity of sweat glands and is composed of tonic

[1] Corresponding Author: Sriram Kumar P, E-mail: psrirmakumar.rs.bme21@iitbhu.ac.in.

and phasic components [1]. The tonic component is a slowly varying or low-frequency signal of EDA. It is influenced by the thermoregulation of the body as well as the surrounding air's humidity and temperature. Additionally, the tonic component includes information on an individual's degree of overall arousal. On the other hand, the phasic component reflects neural stimulation from the sympathetic nervous system and is a fast-varying or high-frequency component of EDA [2]. The performance of emotion detection highly relies on decomposition methods, so it's essential to find a reliable decomposition technique that will improve the human emotion monitoring system. Researchers have proposed a variety of EDA decomposition methods such as non-negative deconvolution, dynamic causal modelling, cubic-spline-based non-negative sparse deconvolution (cvxEDA), compressed sensing, non-negative sparse deconvolution (SparsEDA) [3] and BayesianEDA [4]. The performance of the decomposition methods can be evaluated using feature extraction methods and machine learning algorithms. Time, frequency, and time-frequency domain features calculated from the EDA were used to characterize the emotional states [5]. Linear, non-linear, ensemble, and deep learning-based classifiers were used in the literature for recognizing emotions using EDA signals [6]. In this study, we decomposed the EDA signals using cvxEDA and BayesianEDA methods and calculated the time domain features. The performance of the decomposition methods was evaluated using a statistical significance test and machine learning classifiers such as LR, and SVM.

2. Materials and Methods

The proposed process pipeline followed in this study is shown in Figure 1.

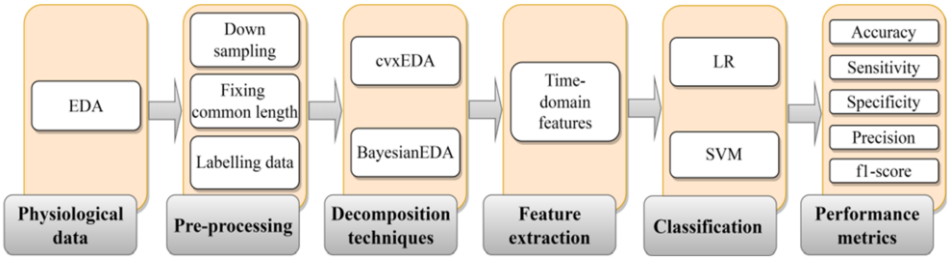


Figure 1. Propose process pipeline

Initially, the EDA signals considered in our study were obtained from the publicly available Continuously Annotated Signals of Emotion (CASE) dataset [7]. The dataset includes recordings of continuously self-annotated physiological signals from 30 participants aged 22 to 37 years (15 male and 15 female). The participants watched eight video clips (mean duration of 158.75 ± 23.67 seconds) to elicit four emotions (amusing, boring, relaxing, and scary, with two videos for each emotion) on a desktop monitor. The video clips were played in different sequences between participants to elicit the appropriate emotions and were recorded in a confined laboratory setting. In the second stage, the EDA signals were down-sampled to 20Hz, and then a common length of 2374 samples was segmented from the end of each EDA signal for the eight video clips (amusing 1 and 2, boring 1 and 2, relaxing 1 and 2, scary 1 and 2) to avoid biasing during the feature extraction process. This optimal length was chosen based on the minimum number of samples available in the dataset for a specific EDA signal

(boring1 has a total length of 2374 samples). All EDA signals of the participants recorded during the amusing1 and amusing2 stimuli were grouped under a single class label 'amusing'. The same was applied to the other three emotions, and the corresponding class labels were 'boring', 'relaxing', and 'scary'. In the third stage, the pre-processed EDA was decomposed into tonic and phasic components using the cvxEDA [8] and BayesianEDA [5] methods. A total of 12 time-domain features (input variables) were extracted from the phasic component of each decomposition technique and normalized from 0 to 1 in the fourth stage, as listed in Table 1 [5], [9].

Table 1. Time domain features

Mean (**MN**), Median (**MDN**), Standard deviation (**STD**), Skewness (**SKW**), Kur tosis (**KRT**), Mean of first derivative (**MFD**), Mean of second derivative (**MSD**), Standard deviation of first derivative (**SFD**), Standard deviation of second deriva- tive (**SSD**), Hjorth complexity (**HC**), Hjorth mobility (**HM**), Hjorth activity (**HA**).

We implemented a non-parametric Wilcoxon rank sum test on the features to check their statistical significance. Furthermore, we fed the features to machine learning methods such as LR and SVM [10] to classify categorical emotions such as amusing, boring, relaxing, and scary (output variables). The models were evaluated using 10-fold cross-validation, and the data were balanced during both the training and test phases, ensuring the same number of observations for each class. We performed machine learning using Python 3.6 and the sci-kit learn packages. Finally, we evaluated the performance of the machine learning models using measures such as accuracy, sensitivity, specificity, precision, and f1-score.

3. Results and Discussions

The representative EDA signals of emotions such as amusing, boring, relaxing, and scary for participants and the corresponding phasic component of all emotions using cvxEDA and BayesianEDA were shown in Figure 2(A)-(C).

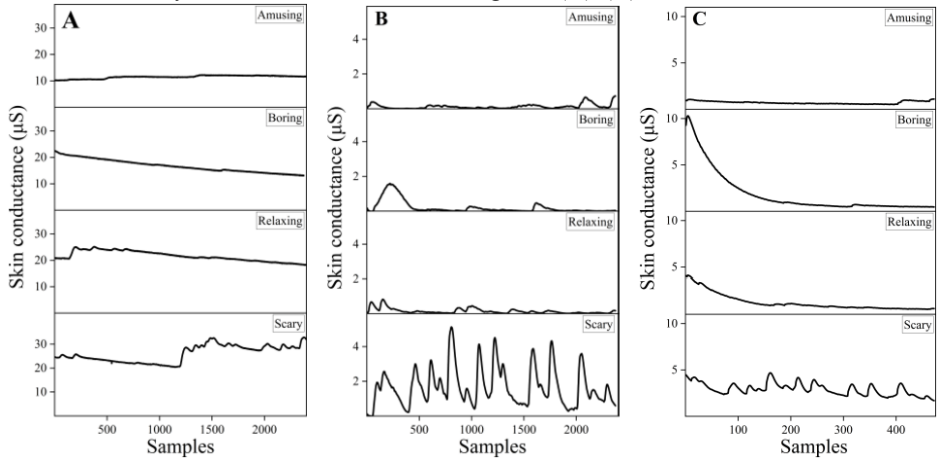


Figure 2. Representative signals of a participant in various emotional states (A) EDA before decomposition, (B) Phasic component deconvolved by cvxEDA, and (C) Phasic component deconvolved by BayesianEDA.

Table 2 shows the statistical significance values of features obtained by the Wilcoxon rank sum test. The six features, such as MN, MDN, STD, SFD, SSD, and HA in cvxEDA were significant for five emotional pairs, and the hypothetical rejection failed

on all features for the boring vs. relaxing emotional pair. In contrast, four features such as STD, MFD, MSD, and HA were significant ($p<0.05$) in BayesianEDA for boring vs. relaxing emotional pairs, and the MFD feature was significant for all emotional pairs. Scary emotion discriminated against the other emotions in most of the features, and it may be due to higher skin conductance caused by more sweat secretion than the other three emotions.

Table 2. Statistical significance of features of cvxEDA and BayesianEDA

	cvxEDA						BayscianEDA					
Feature	AvB	AvR	AvS	BvR	BvS	RvS	AvB	AvR	AvS	BvR	BvS	RvS
MN	*	*	***	0.72	***	***	*	0.54	***	0.13	**	***
MDN	*	*	***	0.98	***	***	0.77	0.58	***	0.82	***	***
STD	**	*	***	0.45	***	***	***	***	***	*	0.20	0.14
SKW	0.35	0.73	0.14	0.78	*	0.10	***	***	***	0.96	***	***
KRT	0.26	0.91	0.36	0.32	0.07	0.45	***	***	0.10	0.77	***	***
MFD	*	0.09	***	0.34	***	***	***	***	***	*	***	*
MSD	0.86	0.60	***	0.30	***	***	**	***	*	***	***	0.21
SFD	*	*	***	0.69	***	***	0.16	0.83	***	0.07	***	***
SSD	*	0.06	***	0.89	***	***	0.64	0.17	***	0.25	***	***
HC	0.54	0.94	*	0.67	*	*	***	***	0.90	0.72	***	***
HM	**	0.05	**	0.16	***	***	***	***	*	0.16	***	***
HA	**	*	***	0.45	***	***	***	***	***	*	0.20	0.14

$^{*}\ p < 0.05,\ ^{**}\ p < 0.005,\ ^{***}\ p < 0.0005$

We fed the features obtained by two decomposition techniques to the classifiers, namely LR and SVM. The average 10-fold cross-validation performance of the two classifiers for four emotions is shown in Figure 3. The results reveal that SVM had greater classification accuracy when employing features derived from the phasic component deconvolved by BayseianEDA. The average classification accuracy, sensitivity, specificity, precision, and f1-score were 88.12%, 76.25%, 92.08%, 76.16%, and 76.15% respectively.

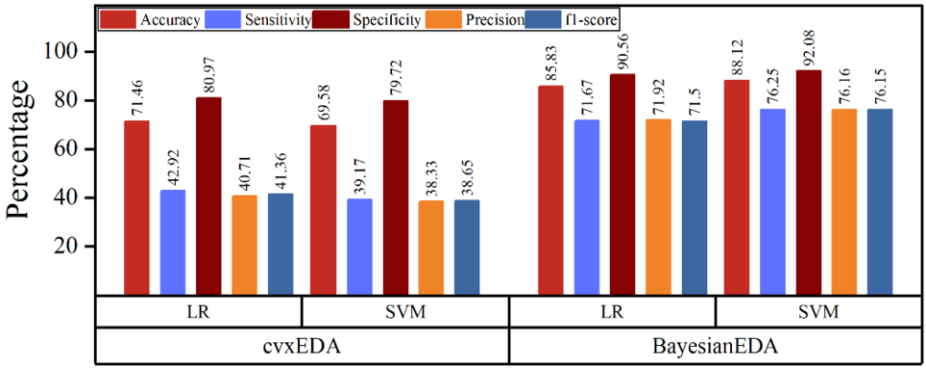


Figure 3. Classification results of LR and SVM classifier

4. Limitations and Future study

In this study, we compared the performance of two decomposition methods using time-domain features and machine-learning algorithms. However, many other decomposition methods were reported in the literature for the deconvolution of EDA data. We evaluated the performance of two decomposition techniques using time-domain features and still need to examine their effectiveness using frequency and time-frequency domain features. We tested with linear classifiers like LR and SVM to detect emotions. In addition, the use of parametric and non-parametric machine learning

algorithms, as well as deep learning-based algorithms and unsupervised learning algorithms, may be explored to increase performance. Moreover, by incorporating additional datasets in the future, we can augment our sample size and enhance the statistical power of our analysis, enabling us to test the effectiveness of decomposition techniques with greater accuracy and precision.

5. Conclusion

In this study, the effectiveness of two decomposition methods in emotion detection was analyzed using feature extraction and machine learning. Initially, cvxEDA and Bayesian EDA were used to decompose the pre-processed EDA signals of four emotional states. Each phasic component of the EDA signals was used to extract the time domain features. The effectiveness of decomposition techniques was further validated using statistical tests and machine learning algorithms like LR and SVM classifiers. The MFD feature extracted from the BayesianEDA method discriminated all the considered six emotional pairs with high statistical significance ($p<0.05$). We used a pipeline that included the BayesianEDA phasic component, time domain features, and SVM to achieve an average 10-fold cross-validation accuracy of 88.12%. The proposed framework can be used for the early diagnosis of psychological conditions to identify emotional states.

References

[1] Boucsein W. Principles of electrodermal phenomena. InElectrodermal activity 2012 (pp. 1-86). Springer, Boston, MA.

[2] Amin MR, Faghih RT. Tonic and phasic decomposition of skin conductance data: A generalized-cross-validation-based block coordinate descent approach. In2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019 Jul 23 (pp. 745-749). IEEE.

[3] Posada-Quintero HF, Chon KH. Innovations in electrodermal activity data collection and signal processing: A systematic review. Sensors. 2020 Jan 15;20(2):479.

[4] Amin R, Faghih RT. Physiological characterization of electrodermal activity enables scalable near real-time autonomic nervous system activation inference. PLoS computational biology. 2022 Jul 28;18(7):e1010275.

[5] Shukla J, Barreda-Angeles M, Oliver J, Nandi GC, Puig D. Feature extraction and selection for emotion recognition from electrodermal activity. IEEE Transactions on Affective Computing. 2019 Feb 26;12(4):857-69.

[6] Ganapathy N, Veeranki YR, Swaminathan R. Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features. Expert Systems with Applications. 2020 Nov 30;159:113571.

[7] Sharma K, Castellini C, van den Broek EL, Albu-Schaeffer A, Schwenker F. A dataset of continuous affect annotations and physiological signals for emotion analysis. Scientific data. 2019 Oct 9;6(1):1-3.

[8] Greco A, Valenza G, Lanata A, et al. (2016) cvxEDA: a convex optimization approach to electro-dermal activity processing. IEEE Trans. Biomed. Eng. 63:797–804. https://doi.org/10.1109/TBME. 2015.2474131.

[9] Zangro´niz R, Mart´ınez-Rodrigo A, Pastor JM, Lo´pez MT, Ferna´ndez-Caballero A. Electrodermal activ- ity sensor for classification of calm/distress condition. Sensors. 2017 Oct 12;17(10):2324.

[10] Abirami S, Thomas J, Yuvaraj R, Agastinose Ronickom JF. A Comparative Study on EEG Features for Neonatal Seizure Detection. InBiomedical Signals Based Computer-Aided Diagnosis for Neurological Disorders 2022 (pp. 43-64). Springer, Cham.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230068

Mapping Korean National Health Insurance Claim Codes for Laboratory Test to SNOMED CT

Sumi SUNG[a], Hyeoun-Ae PARK[b,1], Seul Ki PARK[c], Hyesil JUNG[d], Hannah KANG[b,e] and Min Sun LEE[b,f]
[a]*Office of Hospital Information, Seoul National University Hospital, Seoul, Republic of Korea*
[b]*College of Nursing, Seoul National University, Seoul, Republic of Korea*
[c]*Department of Nursing, Daejeon University, Daejeon, Republic of Korea*
[d]*Department of Nursing, Inha University, Incheon, Republic of Korea*
[e]*Kakao Healthcare Corp., Republic of Korea*
[f]*Healthcare Information Standardization Department, Korea Healthcare Information Service, Republic of Korea*
ORCiD ID: Sumi SUNG https://orcid.org/0000-0003-3897-4698 Hyeoun-Ae PARK https://orcid.org/0000-0002-3770-4998 SeulKi PARK https://orcid.org/0000-0001-7441-8010 Hyesil JUNG https://orcid.org/0000-0002-8346-9343 Hannah KANG https://orcid.org/0000-0001-6567-6894 Min Sun LEE https://orcid.org/0000-0002-0120-4174

Abstract. The aim of this study was to map Korean national health insurance claims codes for laboratory tests to SNOMED CT. The mapping source codes were 4,111 claims codes for laboratory test and mapping target codes were the International Edition of SNOMED CT released on July 31, 2020. We used rule-based automated and manual mapping methods. The mapping results were validated by two experts. Out of 4,111 codes, 90.5% were mapped to the concepts of procedure hierarchy in SNOMED CT. Of them, 51.4% of the codes were exactly mapped to SNOMED CT concepts, and 34.8% of the codes were mapped to SNOMED CT concepts as one-to-one mapping.

Keywords. Systematized Nomenclature of Medicine Clinical Terms, National Health Insurance Reimbursement, Semantic interoperability

1. Introduction

As most medical institutions in South Korea adopted hospital information system (HIS), a large amount of clinical data is being collected at the point of care [1]. In addition, a large amount of personal health data is being collected through wearable devices and mobile applications [2]. However, the collected data is not fully utilized due to the lack of data sharing strategies [3]. To use the collected data for patient care or clinical research,

[1] Corresponding Author: Hyeoun-Ae PARK, College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea; E-mail: hapark@snu.ac.kr

it is essential to ensure semantic interoperability through the use of health data standards including standard terminologies.

In 2020, to achieve semantic interoperability in healthcare, South Korea joined the SNOMED International as the 39th member country. As a way to promote the use of SNOMED CT in Korea, there have been various national initiatives such as mapping Korean Classification of Disease-7 (KCD-7), national health insurance claims codes which is also called electronic data interchange (EDI) codes for procedures, pharmaceutical products, narrative medical records of gastrectomy patients, and Korean national health checkup questionnaire to SNOMED CT [4-8]. However, the national health insurance claims codes for laboratory tests were not mapped to SNOMED CT, limiting the use of laboratory tests information for patient care or clinical research.

With this background, the aim of this study was to map national health insurance claims codes for laboratory tests to SNOMED CT to improve semantic interoperability.

2. Methods

The mapping source codes used in this study comprise 4,111 codes for laboratory tests covered by the national health insurance service in 2020. The codes for laboratory tests include codes for evaluation procedures such as complete blood counts, electrolytes, liver function tests, poisoning screening tests and genetic tests.

The mapping target terminology is the International Edition of SNOMED CT released on July 31, 2020. Before we selected SNOMED CT as a target terminology, we also reviewed LOINC as a possible target terminology. When we reviewed insurance claims codes for laboratory tests, most of EDI codes did not have 6 parts of LOINC, namely component, property, time, system, scale, and method. An example is presented in Table 1. We selected SNOMED CT as target terminology over LOINC. The target concepts were restricted to the concepts in the 'Procedure' top-level hierarchy.

Table 1. Comparison of the mapping results between SNOMED CT and LOINC

Mapping sources	Map target	
EDI	**SNOMED CT**	**LOINC**
D1850 ALT(SGPT)	34608000 |Alanine aminotransferase measurement (procedure)|	16324-6 Alanine aminotransferase:CCnc:Pt:RBC:Qn 76625-3 Alanine aminotransferase:CCnc:Pt:Bld:Qn 54492-4 Alanine aminotransferase:CCnc:Pt:Plr fld:Qn 96586-3 Alanine aminotransferase:CCnc:Pt:Bld.dot:Qn 50168-4 Alanine aminotransferase:CCnc:Pt:Dial fld:Qn 1742-6 Alanine aminotransferase:CCnc:Pt:Ser/Plas:Qn 25302-1 Alanine aminotransferase:CCnc:Pt:Body fld:Qn 1741-8 Alanine aminotransferase:CCnc:Pt:Amnio fld:Qn 54491-6 Alanine aminotransferase:CCnc:Pt:Periton fld:Qn 77144-4 Alanine aminotransferase:CCnc:Pt:Ser/Plas/Bld:Qn

Mapping process is consisted of 5 steps: 1) understanding meaning of the source terms, 2) identifying the mapping rules, 3) automatic or manual mapping, 4) classification and 5) validation as presented in Figure 1.

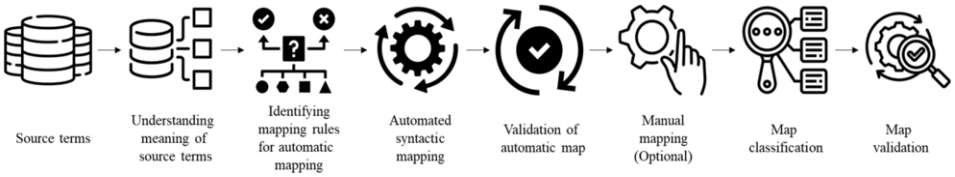


Figure 1. The overall mapping processes.

To understand the meaning of the source terms, we analyzed the laboratory test labels and extracted the component, property, time, system, scale, and method of the laboratory tests and presented them in a structured format.

We identified the pre-processing rules by manually mapping randomly selected 800 source terms to SNOMED CT (as presented in Table 2). Automated syntactic mapping was conducted using self-developed solution. If the automated mapping failed to find the SNOMED CT concept matched, the pre-processing rules were applied according to the numerical order. The automated syntactic mapping results were validated for the semantic match. If the maps did not semantically match or automatic mapping failed, we manually mapped the laboratory test codes to SNOMED CT. The manual mapping prioritized mapping the laboratory test codes to the pre-coordinated concept. If there is no pre-coordinated SNOMED CT concept matching to the source codes, we mapped to post-coordinated expression.

Table 2. Pre-processing mapping rules identified for automated syntactic mapping

Orders	Mapping rules
1	Extract the term within "( )" or the term after "-" in the source term, and then match to SNOMED CT
2	Change pleural to singular words in the source term, and then match to SNOMED CT
3	Remove the following brackets and terms in the source term, and then match to SNOMED CT 1) "( )" and the term within "( )" 2) "[ ]" and the term within "[ ]"
4	Add one of the following terms to the source term, and then match to SNOMED CT 1) Level 2) Measurement 3) Measurement of 4) Test 5) Assay
5	Add one of the following terms to the term after "-" or "_" of source term, and then match to SNOMED CT 1) Level 2) Measurement 3) Measurement of
6	Combine the extracted 6 parts, and then match to SNOMED CT

The maps were classified by expression, map cardinality, and map correlation. Maps were classified as 'pre-coordinated concept' or 'post-coordinated expression' based on expression of target terminology. Maps were classified as 'one to one' or 'one to many' according to the map cardinality. Maps were classified as 'exactly mapped,' 'broadly mapped,' 'partially mapped,' or 'not mapped' according to the map correlation. The examples were presented in Table 3.

Table 3. The examples of the maps

Mapping methods	Source codes	Target SNOMED CT	Classification		
Automatic mapping	D2280 Creatinine	70901006 |Creatinine measurement (procedure)|	Pre-coordinated concept	One to one	Exactly mapped
Automatic mapping	D3061 Hemoglobin A1c	43396009|Hemoglobin A1c measurement (procedure)|	Pre-coordinated concept	One to one	Exactly mapped
Automatic + manual mapping (revised)	D0001020 Complete blood cell count- RBC Count [Microscope]	14089001|Red blood cell count (procedure)|:424226004|Using device (attribute) | = 65473004|Microscope, device (physical object)|	Post-coordinated expression	One to one	Exactly mapped
Automatic + manual mapping (revised)	D1071 D-dimer, Qualitative [Immunoassay]	70648006|D-dimer assay (procedure)| + 414464004|Immunoassay method (procedure)| :370132008|Scale type (attribute) | = 26716007|Qualitative (qualifier value)|	Post-coordinated expression	One to many	Exactly mapped
Manual mapping	Hemolytic Anemia- Auto Hemolysis Test	401297005|Hemolysis screening test (procedure)|	Pre-coordinated concept	One to one	Broadly mapped

The maps were finally validated by two experts who have experiences in SNOMED CT mapping. When the two experts did not agree on the map, the maps were discussed in group meetings attended by the project manager and research team members who were not involved in the mapping process.

3. Results

In automatic mapping, 3,406 (82.9%) of the total codes were mapped correctly, 1,764 (42.9%) were applied to pre-processing rules 1 to 5, and 1,642 (39.9%) were applied to rule 6, which matches SNOMED CT by combining structured 6 parts. 191 (4.6%) were mapped incorrectly and 514 (12.5%) were not mapped. Incorrectly mapped or unmapped were revised manually by the authors.

Of the 4,111 source codes, 1,574 (38.2%) were mapped to pre-coordinated concept, 2,148 (52.2%) were mapped to post-coordinated expression. 1,434 (34.8%) were mapped to one concept, and 2,288 (55.6%) were mapped to more than two concepts of procedure hierarchy. 2,114 (51.4%) were exactly mapped, 1,505 (36.5%) were broadly mapped, and 103 (2.5%) were partially mapped. 385 (9.5%) codes were not mapped to SNOMED CT concept. The final map is publically available on the website of Healthcare Information Standard by the Korea Health Information Service [9].

4. Discussion

To the best of our knowledge, this study is the first attempt to map the Korean health insurance claims codes for laboratory tests to SNOMED CT rather than LOINC. This study attempted to identify mapping rules and map the claims codes using the automatic

mapping method. If automatic mapping failed, the source codes were manually mapped by combining structured 6 parts after analyzing the source codes. As a result, a total of 90.1% of EDI codes were mapped to SNOMED CT, and among them, 39.9% were mapped by combining 6 parts of the source codes which implied that structuring laboratory test codes will improve the mapping rate to internationally standardized laboratory terms. In addition, using SNOMED CT instead of LOINC as the map target, it is possible to qualify details of laboratory tests other than 6 parts of LOINC in more detail with SNOMED CT attribute-value pairs. Examples of SNOMED CT attributes include '363702006 |Has focus (attribute)|', '363703001 |Has intent (attribute)|', '424226004 |Using device (attribute)|', and '260507000 |Access (attribute)|'.

The map developed in this study can be used for mapping codes or terms for laboratory tests used in Korean medical institutions to SNOMED CT in the future. Since codes used in the hospitals in Korea are linked to the EDI code for national health insurance reimbursement, refining the map developed in this study with more attribute-value pairs will be useful for introducing SNOMED CT in hospitals.

5. Conclusion

In this study, we developed the map between Korean national health insurance reimbursement claims codes for laboratory test and SNOMED CT throughout automatic and manual mapping methods. The map we developed can be used to map local codes for laboratory tests to SNOMED CT in the local hospital. Finally, the map will facilitate semantic interoperability at the point of care and use for clinical research.

Acknowledgements

This research was supported by the Korea Health Information Service.

References

[1] Park YT. The current status and future plan for electronic medical record system at medical institution in South Korea. HIRA Research. 2017:11(2):52-61
[2] Schwab K. The global competitiveness report 2019. World Economic Forum. 2019.
[3] Park HA, Yu SJ, Jung H. Strategies for Adopting and Implementing SNOMED CT in Korea. Healthc Inform Res. 2021 Jan;27(1):3-10. doi: 10.4258/hir.2021.27.1.3.
[4] Korea Health Information Service. KCD-SNOMED CT Mapping Table 2022 [cited April 6 2022]. Available from: https://www.hins.or.kr/termMapping/kcdMappingList.es?mid=a11301020000.
[5] Jung H, et al. Mapping Korean National Health Insurance Pharmaceutical Claim Codes to SNOMED CT. Stud Health Technol Inform. 2022 May 25;294:297-301. doi: 10.3233/SHTI220462.
[6] Kang H, Park HA. Mapping Korean National Health Insurance Reimbursement Claim Codes for Therapeutic and Surgical Procedures to SNOMED-CT to Facilitate Data Reuse. Stud Health Technol Inform. 2022 Jun 6;290:101-105. doi: 10.3233/SHTI220040.
[7] So EY, Park HA. Exploring the Possibility of Information Sharing between the Medical and Nursing Domains by Mapping Medical Records to SNOMED CT and ICNP. Healthc Inform Res. 2011;17(3):156-61. doi: 10.4258/hir.2011.17.3.156.
[8] Hwang JE, Park HA, Shin SY. Mapping the Korean National Health Checkup Questionnaire to Standard Terminologies. Healthc Inform Res. 2021;27(4):287-297. doi: 10.4258/hir.2021.27.4.287.
[9] Korea Health Information Service. EDI-SNOMED CT Mapping Table 2022 [cited March 3 2023]. Available from: https://www.hins.or.kr/termMapping/ediMappingList.es?mid=a11301030000.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230069

WHO's Community-Centered Epidemic and Pandemic Information Platform: Hive

Sylvie BRIAND [a], Sarah HESS [a,1], John LEE [a,2], Katherine SHERIDAN [a], Tim NGUYEN [a] and Brian YAU [a]
[a] *Department of Epidemic & Pandemic Preparedness & Prevention, Health Emergencies Programme, World Health Organization, Geneva, Switzerland*

Abstract. Each epidemic and pandemic is accompanied by an infodemic. The infodemic during the COVID-19 pandemic was unprecedented. Accessing accurate information was difficult and misinformation harmed the pandemic response, the health of individuals and trust in science, governments and societies. WHO is building a community-centered information platform, the Hive, to deliver on the vision of ensuring that all people everywhere have access to the right information, at the right time, in the right format in order to make decisions to protect their health and the health of others. The platform provides access to credible information, a safe space for knowledge-sharing, discussion, and collaborating with others, and a forum to crowdsource solutions to problems. The platform is equipped with many collaboration features, including instant chats, event management, and data analytics tools to generate insights. The Hive platform is an innovative minimum viable product (MVP) that seeks to leverage the complex information ecosystem and the invaluable role communities play to share and access trustworthy health information during epidemics and pandemics.

Keywords. Trust, trustworthy information, powered by communities, high-impact health events preparedness, epidemic and pandemic response, health information platform, machine-learning

1. Introduction

In today's highly interconnected world, each individual is exposed to a complex information ecosystem that spans both the physical and digital environments. An *infodemic,* is an overabundance of information, accurate or not, in the digital and physical space, accompanying an acute health event such as an outbreak or epidemic [1]. During epidemics and pandemics there are changes to the information ecosystem due to the infodemic and changes to the information-seeking behaviors of individuals, due to an increase in uncertainty, the evolving situation and the need to make decisions rapidly. The infodemic that accompanied the COVID-19 pandemic presents an unprecedented example. There was an increase in volume of scientific information with more than 20,000 COVID-19 related articles published in the first six months of the pandemic [2]. Many of the articles were generated rapidly without adequate scientific rigor, and peer review. Furthermore, scientific knowledge evolved as the pandemic evolved and this

[1] Corresponding Author: Sarah Hess, Department of Epidemic & Pandemic Preparedness & Prevention, Health Emergencies Programme, World Health Organization, Geneva, Switzerland, E-mail: hesss@who.int

[2] Corresponding Author: John Lee, Department of Epidemic & Pandemic Preparedness & Prevention, Health Emergencies Programme, World Health Organization, Geneva, Switzerland, E-mail: leesung@who.int

evolution was often not explained well. There were many voices and opinions debating the science and accompanying policies, which often led to polarizing debates. In some cases, response measures were politicized and media content sensationalized, which amplified uncertainty and fear within societies. Under these circumstances, it was difficult for people to access trustworthy information in order to make decisions to protect their health and the health of their communities. All of these aspects of the infodemic impacted trust within societies, including trust in institutions, experts, health authorities, governments, and academia, thereby negatively impacting pandemic response efforts. The global health community looked to find solutions that improve access to trustworthy information, increase health and scientific literacy, and facilitate individual and community decision-making while promoting autonomy and localization of epidemic and pandemic responses.

2. Trustworthy Information Powered by Communities – "Caring is Sharing"

The scale and unpredictability of the COVID-19 pandemic placed extraordinary challenges on leaders, decision-makers, and communities. Communities and individuals need to be listened to and able to participate in broader discussions to share and understand the challenges and impediments they are facing together. If questions and concerns are not addressed adequately and information voids persist, then rumors and misinformation can flourish. For example, suppose Public Health and Social Measures being mandated by authorities are not feasible to implement in certain settings or require resources that are not available. In that case, there can be a disconnect between leadership and communities and individuals – leading to a breakdown in trust. Trust as a social capital, is fragile particularly in times of crisis. It requires specific, intentional interventions for it to be nurtured and maintained. Autonomy, consistency, and transparency in communication underpinned by the commitment and accountability of leaders and decision-makers and active community involvement are essential to developing and maintaining trust across our communities when the topic concerns lives, livelihoods, health, and well-being. Solutions are needed that build trust, are inclusive and responsive, and enable autonomy.

3. A Solution – The Hive Platform

3.1. A Space for Communities to Connect and Share Best Practice

WHO uses several strategies to disseminate information to the public and decision-makers. The WHO Information Network for Epidemics ("EPI-WIN") [3] is one strategy that provides resources and regular updates using a whole-of-society approach. Recognizing and promoting local knowledge and expertise is critical to ensure WHO's guidance or public-health interventions are relevant, feasible and appropriate. During the COVID-19 pandemic, a wealth of information and experience was shared between industries, decision-makers and within communities. However, it was often ad-hoc and opportunistic. WHO wants to facilitate systematic sharing of this local knowledge and expertise, capitalize on best-practice, and support decision-makers and community leaders to connect and share experiences to prepare and respond to high-impact health events.

The Hive, WHO's community-centered Epidemic and Pandemic Information Platform [4], is designed to complement the EPI-WIN approach and enhance the way that WHO supports and learns from communities. Communities are where trust is built, information is shared and collaboration happens. Particularly in times of crisis and uncertainty, people turn to those who have remained trustworthy over time. "For individuals to adopt, change and sustain new behaviors during epidemics, they need to… have the ability to enact the recommendations in their living/social/work/faith setting" [5]. The Hive platform is designed to bring together communities, while leveraging current technology and the digital information ecosystem. The Hive platform, like a beehive, has the opportunity to be space of activity, support and community.

3.2. The Hive Platform as a System

To deliver on the vision of ensuring that all people everywhere have access to the right information, at the right time and in the right format, the platform must provide access to credible and trustworthy information. Given the complex nature of the current information ecosystem, and how individuals seek, engage and share information, it would be an impossible task for WHO alone. As depicted in Figure 1. the Hive as a system is designed to leverage the power of communities and the opportunities to be gained from community participation. Moreover, the data gathered from its feedback loops will guide the machine learning functionality of the Hive, presenting personalized and appropriate information relevant to the interests of the individual. We envision the Hive platform enabling an inclusive response to public health events through multisectoral, multi-level communication and collaboration. Future development will include scaling up and rolling out to global communities; gamification to encourage community participation; advanced analytics and integrated listening tools to guide community engagement activities, support early detection and targeted response measures. Hive will place the community and the community's information needs at the center and provide the tools necessary for knowledge exchange and co-creation.

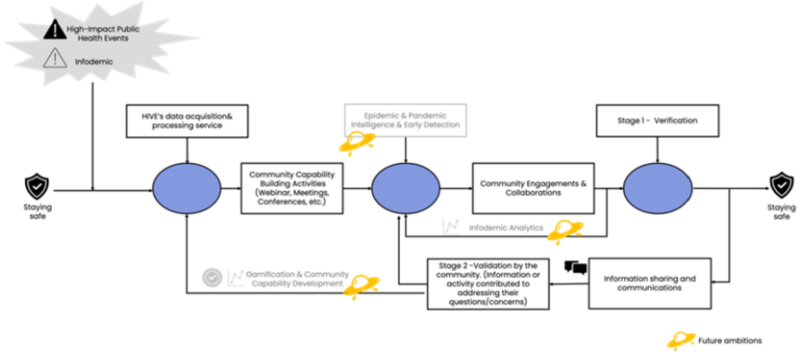


Figure 1. The Hive Platform as a System

3.3. The Technology Behind the Hive – A Process of Evolution and Refinement

The Hive system uses a continuously improving data acquisition system that gathers information from a defined list of global sources. The data acquisition service feature gathers the information from external sources, such as mainstream news outlets, and

health authority websites, using web crawling and scraping. The process is triggered regularly to ensure up-to-date information on the Hive platform. After crawling and scraping the information, the content and metadata are stored and accessed by the indexer for enrichment and indexing. Then the system tags each content and transforms texts into numerical values for machine learning.

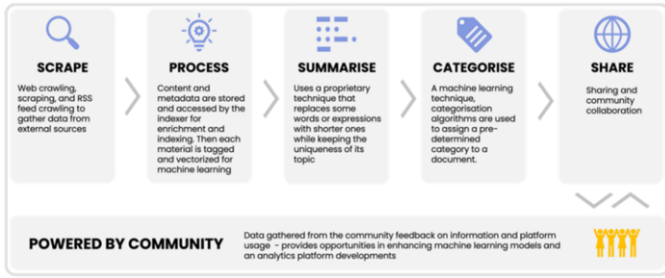


Figure 2. Data acquisition and data processing features of the HIVE

Hive's summarization feature provides a short summary of the presented information. This process uses a technique that replaces some words or expressions with shorter ones while keeping the uniqueness of its topic. A machine learning technique, categorization algorithms assign a pre-determined category to each content. Since the Hive system has been recently launched, it is currently using an unsupervised model. After onboarding early adaptors, the system should be able to receive feedback to assess and improve the classification performance. The data gathered from community feedback and usage patterns will guide machine learning models to produce personalized and appropriate information relevant to the interests of the individual. These multistep processes are depicted in Figure 2.

3.4. Identifying & Verifying Trustworthy Sources of Information

The Hive uses three main pathways to gather information from all online content and electronic media: 1) web crawling and scraping. 2) co-created content – jointly produced information between combinations of industry, research, government, and civil society. 3) the community members "suggest" sources that can be added to the platform.

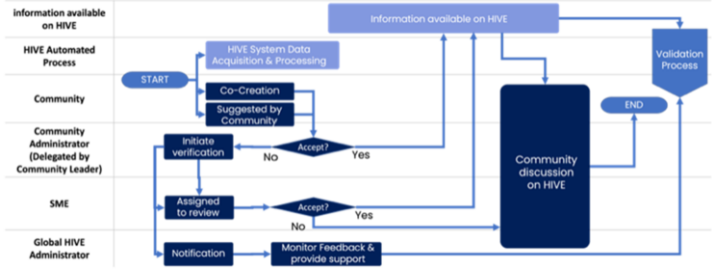


Figure 3. Verification Process

Figure 3. shows the workflow of integrating the inputs from Subject Mater Experts (SMEs) into the initial verification of information that can be included in the Hive information universe. This approach allows the communities to access large amounts of information from online media in a digestible way while generating and adding tailored information to the Hive platform through the review and acceptance process.

3.5. Process of Validating the Information

One of the platform's key attributes is that the information can address questions and concerns of the community. That is, the community's validation process can provide essential data for measuring the effectiveness of measures used to build trust during epidemics and pandemics. Figure 4. illustrates the workflow of validation by the community and the data capture workflow for measurements of the relevance of information on Hive.

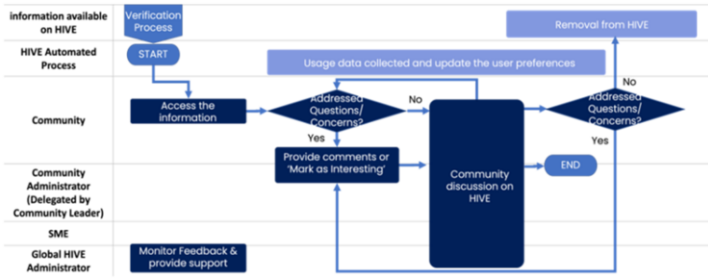


Figure 4. Validation Process

Quantitative and qualitative feedback on whether the content has contributed to addressing their questions and concerns and usage data can provide insights into the latest trend and best approaches.

4. Conclusion

The Hive is an ambitious, future-facing platform designed to transform the ways of working with communities during public health emergencies. It is an innovation that will support developing and maintaining community trust, enabling active community collaboration to address questions and concerns, and share trustworthy and relevant information. In future work, we aim to continuously improve the ways of working for the communities on Hive and bringing in the latest information technology advancements, including data analytics and machine learning to expand the role of the community in high-impact public health events and complement Epidemic & Pandemic Preparedness & Prevention.

References

[1] World Health Organization Infodemic. Available at: https://www.who.int/health-topics/infodemic#tab=tab_1 Accessed 3 Jan 2023
[2] Teixeira da Silva JA, Tsigaris P & Erfanmanesh M (2021) Publishing volumes in major databases related to Covid-19. Scientometrics 126(1):831–842. https://doi.org/10.1007/s11192-020-03675-3
[3] WHO, EPI-WIN: WHO Information Network for Epidemics. Available at: https://www.who.int/teams/epi-win. Accessed 3 Jan 2023
[4] WHO Hive platform, community centred health information platform. Available at https://hivecenter.who.int/. Accessed 3 Jan 2023
[5] Sylvie Briand, Sarah Hess, Tim Nguyen, Elisabeth Wilhelm , Tina Purnat Managing an Infodemic, Infodemic Management. In press 2023

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230070

Automatic Outlier Detection in Laboratory Result Distributions Within a Real World Data Network

Aída MUÑOZ MONJAS[a1], David RUBIO RUIZ[a,b], David PÉREZ-REY[a] and Matvey PALCHUK[b]
[a] *Biomedical Informatics Group, Universidad Politécnica de Madrid, Spain*
[b] *TriNetX , LLC, Cambridge, MA, USA*
ORCiD ID: Aída Muñoz Monjas https://orcid.org/0000-0003-4504-9775 , David Rubio Ruiz https://orcid.org/0000-0003-1880-8019, David Perez-Rey https://orcid.org/0000-0002-9021-2597, Matvey Palchuk https://orcid.org/0000-0002-7737-8752

Abstract. Laboratory data must be interoperable to be able to accurately compare the results of a lab test between healthcare organizations. To achieve this, terminologies like LOINC (Logical Observation Identifiers, Names and Codes) provide unique identification codes for laboratory tests. Once standardized, the numeric results of laboratory tests can be aggregated and represented in histograms. Due to the characteristics of Real World Data (RWD), outliers and abnormal values are common, but these cases should be treated as exceptions, excluding them from possible analysis. The proposed work analyses two methods capable of automating the selection of histogram limits to sanitize the generated lab test result distributions, Tukey's box-plot method and a "Distance to Density" approach, within the TriNetX Real World Data Network. The generated limits using clinical RWD are generally wider for Tukey's method and narrower for the second method, both greatly dependent on the values used for the algorithm's parameters.

Keywords. Outlier detection, laboratory test, real world data, LOINC, interoperability

1. Introduction

Over the last decades, healthcare systems have been undergoing a digitalization with significant implications for primary and secondary uses of clinical data. During this process, Health Care Organizations (HCOs) have mainly started storing their using Electronic Health Records (EHRs). The correct handling and processing of this data is essential, not only to guarantee patient safety, but also if these information sources are to be used for secondary purposes such as research [1].

In these institutions, laboratory data is commonly stored using local terminologies, due to the adaptability they enable. Associating local codes to standardized codes in terminologies like LOINC (Logical Observation Identifiers, Names and Codes) improves the potential uses of this data. To facilitate the study and understanding of laboratory tests, their results can be graphically represented. In quantitative laboratory tests where

[1] Corresponding Author: Aída Muñoz Monjas, e-mail: aida.munozm@alumnos.upm,es

the result is a number and a unit, the data can be plotted on histograms that represent the volume of results (on the y-axis) over the numeric value (on the x-axis). These histograms can have different shapes according to the nature of the test and specific characteristics of the population it is performed on.

In general, laboratory test results follow a Weibull distribution, which is defined by its two parameters: shape (k) and scale (λ) [2]. If k=5 and λ =1, the generated distribution approximates to a normal distribution, while for k=0.5 and λ =1 it approximates to an exponential distribution.

In a normal distribution, the expected result of the test is the number the distribution is centred on. Characteristics like mean, standard deviation, skewness and kurtosis are maintained throughout HCOs for each laboratory test, with positive skewness as the most common attribute of this type of distributions. In an exponential distribution, the expected result of the test is usually 0, an absence of the tested substance, having exponentially fewer positive results on the right tail of the distribution.

One of the characteristics of Real World Data (RWD) is the existence of outliers, observations with a great deviation from the rest of the observations registered [3] that add no value to the dataset. Outliers in a clinical context can be due to errors in the data registration, where the physician inputs the value with a unit that is not expected by the system, for example 500 g, but the data is represented with the same numeric value but different unit, 500 kg, in the distribution. Outliers can also be due to errors in the device that performs the lab test, resulting in negative amounts of substance measured, -100 mg/dL of glucose in blood, or physically impossible results, 10^{12} mg/dL of glucose in blood.

The main objective of this work is to automatically eliminate these outliers from laboratory result distributions to facilitate the understanding, visualization, and analysis of the obtained results.

2. Methods

Outlier detection in a laboratory result distribution brings the focus to the relevant information by defining a set of limits that exclude the majority of these inconsistent values. Several methods have been used in the past to define these limits [3]. In this work, two different methods are compared: Tukey's box-plot method, one of the most used approaches, and a "Distance to Density" method, proposed by Last et al. [4].

Tukey's method for outlier detection flags a value located between the inner and outer limits as a possible outlier, while a value outside the outer limit is a clear outlier of the distribution [5]. The inner and outer limits are calculated according to Eq. (1),

$$inner\ limits = \begin{cases} Q3 + 1.5 \cdot IQR \\ Q1 - 1.5 \cdot IQR \end{cases} \qquad outer\ limits = \begin{cases} Q3 + 3 \cdot IQR \\ Q1 - 3 \cdot IQR \end{cases} \tag{1}$$

where $Q1$ and $Q3$ are the first and third quartiles respectively, and IQR is the interquartile range calculated as $IQR = Q3 - Q1$. Tukey's method is useful for skewed distributions, as it does not depend on the mean or standard deviation of the distribution [3].

The Distance to Density method [4] defines the reliability of each data element based on its distance to the values nearby and the weight of its frequency on the distribution. Reliabilities close to 0 characterize outliers, that will be filtered according to a threshold α [4]. Reliability is calculated from above (μ_{RL}) and below (μ_{RH}), according to Eq. (2).

$$\mu_{RLj} = \frac{2}{1+e^{\frac{\left(\beta \cdot M \cdot D \cdot \left(V_{j+1} - V_j\right)\right)}{\left(N_j \cdot \left(V_{j+M+1} - V_{j+1}\right)\right)}}} \qquad \mu_{RHj} = \frac{2}{1+e^{\frac{\left(\beta \cdot M \cdot D \cdot \left(V_j - V_{j-1}\right)\right)}{\left(N_j \cdot \left(V_{j-1} - V_{j-M-1}\right)\right)}}} \quad (2)$$

where D is the total number of observations, V_j is the value of index j with frequency N_j, β is the shape factor representing the attitude towards the distance between succeeding values and M is the lookahead. This method relies heavily on the values of parameters β, M and α, which should be adjusted to fit the results to the desired usecase.

3. Results

In order to assess the quality of the distribution limits generated by these two methods in RWD, limits were generated for the result distributions of sixteen laboratory LOINC codes. The limits generated by both methods were compared against a list of manually curated limits by an expert based on recent literature. Each histogram depicts the numeric results obtained in a laboratory test versus their frequency, grouping the results into bins. The histograms represent the studied data without modifications, which makes distributions difficult to analyse in certain cases, such as code 2744-1 in Figure 1, when outliers are present.

The processed data was obtained from TriNetX's global research network [6], by aggregating the data of EHRs from over 110 million patients across 125 HCOs. The histograms studied in this work are not raw data from a single organization, but the aggregation of multiple sources, ensuring the protection of clinical data and meeting the legal requirements for its handling, i.e., HIPAA, GDPR, etc. [7].

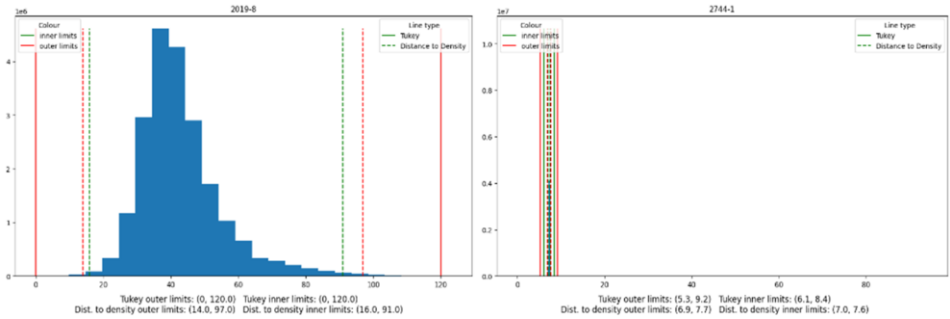


Figure 1. Visualization of generated limits for LOINC codes 2019-8 (left) and 2744-1 (right).

Table 1 contains the obtained results for a group of LOINC codes and their reference values. These reference values depict highest or lower recorded measures or values that are not physiologically possible, or represent manually curated limits generated by an expert. A very small percentage of patients are expected to have lab values outside these limits, and they can be considered good reference values for the exclusion of outliers in each of these tests.

In order to compute these limits, Tukey's method was applied as explained, and two different values for α were used in the Distance to Density method to obtain the inner and outer limits. The parameter M (lookahead) took a value of 1% of the data values, while the inner and outer α (thresholds) were $0.75 \cdot 10^{-4}$ and $0.75 \cdot 10^{-8}$ respectively,

with a β of 0.01. The generated limits were plotted against the distribution to visually assess their quality, as seen in Figure 1.

These inner and outer limits are used in TriNetX to define the plotting area of the distribution (inner limits) and to select the data that will be included when calculating measures such as average mean and standard deviation (outer limits).

Table 1. Obtained inner and outer limits with each method for seven LOINC codes. Comparison with the reference values and manually curated limits

LOINC code	Method	Inner limits		Outer limits	
		generated	reference	generated	reference
14749-6: Glucose in Ser/Pl. (mmol/L)	tukey	(0, 27)	(2.3 [8], 30)	(0, 39)	(0.7 [8], 40)
	distDen	(3.3, 8.3)		(2.8, 9.9)	
2019-8: CO2 in Art.Bld (mm[Hg])	tukey	(0, 120)	(8, 115[9][10])	(14, 97)	(0, 240)
	distDen	(16, 91)		(2.8, 9.9)	
33959-8: Procalcitonin in Ser/Pl. (ng/mL)	tukey	(0, 150)	(0, 9.7 [11])	(0, 150)	(0, 20)
	distDen	(0, 5.1)		(0, 7.2)	
8302-2: Body Height ([in us])	tukey	(0, 110)	(15, 90) ((38.1, 228.6) cm)	(0, 150)	(0, 107.09 [12]) ((0, 272) cm)
	distDen	(47, 76)		(33, 78)	
26881-3: Interleukin-6 in Ser/Pl. (pg/mL)	tukey	(0, 740)	(0, 800)	(0, 1200)	(0, 2221 [13])
	distDen	(0, 1200)		(0, 7600)	
2744-1: PH of Arterial Blood	tukey	(6.1, 8.4)	(4, 9)	(5.3, 9.2)	(0, 9)
	distDen	(7, 7.6)		(6.9, 7.7)	
12841-3: Prostate Specific Ag free/total in Ser/Pl. (%)	tukey	(0, 88)	(0, 80)	(0, 88)	(0, 100)
	distDen	(2, 67)		(0.87, 67)	

The limits for these sixteen LOINC codes were generated using both methods. Tukey's outer limits included over 95% of observations in every code, while the outer limits generated by the Distance to Density method included the same percentage in 13 out of the 16 codes. Both methods included over 80% of the observations in every code between the outer limits, and over 75% between the inner limits.

4. Discussion

Both methods studied in this work managed to shorten the interval of relevant information for the tested LOINC codes, generating limits that can separate outliers from the rest of the distribution in current and future data.

The limits generated by Tukey's method are generally wider than those manually generated by experts in the field, but provide an acceptable approximation to considerably improve the analysis and visualization of these distributions.

The Distance to Density method performs correctly with normal distributions, while being more unreliable with exponential distributions. This is likely due to the values of the algorithm's parameters, which were mainly selected to fit normal distributions, as exponential distributions limits are not clearly defined concerning outliers. When comparing the limits generated by this method with the limits generated by Tukey's method, it can be observed that these limits are closer to the distribution peaks, while Tukey's limits are generally wider.

The main limitation of both algorithms is the appropriate selection of the values of their parameters, as these greatly affect the obtained results. Generating limits that are too narrow can exclude relevant information, while generating overly wide limits produces a less readable graph and increases the complexity of the histogram representation.

5. Conclusions

Outliers are inherent to RWD, but still allow data analysis. The detection and exclusion of outliers cleans up the numeric results obtained in laboratory distributions and facilitates the analysis and interpretation of the result histograms. The main aim of this work is to provide a comparison between the results obtained by Tukey's approach to outlier detection and the "Distance to Density" method in clinical RWD. Tukey's method proved to generate limits wider than those annotated by an expert, while the Distance to Density method generated narrower limits. Both methods produced an acceptable approximation to improve the analysis and visualization of the distributions, but additional manual curation is recommended for optimal results.

References

[1] Miriovsky B, Shulman L, Abernethy A. Importance of Health Information Technology, Electronic Health Records, and Continuously Aggregating Data to Comparative Effectiveness Research and Learning Health Care. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2012 10;30.

[2] Hallinan AJ Jr. A Review of the Weibull Distribution. Journal of Quality Technology. 1993;25(2):85-93. Available from: https://doi.org/10.1080/00224065.1993.11979431.

[3] Seo S. A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets. 2006 Aug. Available from: https://www.semanticscholar. org/paper/A-Review-and-Comparison-of-Methods-for-Detecting-in-Seo/ cb868f0b242b9623b7544a58b6a21647dfa138a5.

[4] Last M, Kandel A. Automated Detection of Outliers in Real-World Data. 2002 03.

[5] Dawson R. How Significant is a Boxplot Outlier? Journal of Statistics Education. 2011;19:2. Available from: https://doi.org/10.1080/10691898.2011.11889610.

[6] Topaloglu U, Palchuk M. Using a Federated Network of Real-World Data to Optimize Clinical Trials Operations. JCO Clinical Cancer Informatics. 2018 02;2:1-10.

[7] TriNetX. Home; 2022. Available from: http://trinetx.com.

[8] Cryer P. Hypoglycemia, functional brain failure, and brain death. The Journal of clinical investigation. 2007 04;117:868-70.

[9] Goila A, Pawar M. The diagnosis of brain death. Indian journal of critical care medicine : peer-reviewed, official publication of Indian Society of Critical Care Medicine. 2009 03;13:7-11.

[10] Wu Z, Luo Z, Luo Z, Ge J, Jin J, Cao Z, et al. Baseline Level and Reduction in PaCO2 are Associated with the Treatment Effect of Long-Term Home Noninvasive Positive Pressure Ventilation in Stable Hypercapnic Patients with COPD: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. International Journal of Chronic Obstructive Pulmonary Disease. 2022 04;Volume 17:719-33.

[11] Rangeard O, Audibert G, Perrier JF, Loos-Ayav C, Lalot JM, Agavriloaie M, et al. Relationship Between Procalcitonin Values and Infection in Brain-Dead Organ Donors. Transplantation Proceedings. 2007;39(10):2970-4. Available from: https://www.sciencedirect.com/science/article/pii/S0041134507009670.

[12] Records GW. Tallest man ever;. https://www.guinnessworldrecords.com/world-records/tallest-man-ever.

[13] Amado J, López-Espadas F, Vázquez-Barquero A, Salas E, Riancho JA, López-Cordovilla JJ, et al. Blood levels of cytokines in brain-dead patients: Relationship with circulating hormones and acute phase reactants. Metabolism. 1995;44(6):812-6. Available from: https://www.sciencedirect.com/science/article/pii/0026049595901981.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230071

The Importance of Being FAIR and FAST – The Clinical Epidemiology and Study Platform of the German Network University Medicine (NUKLEUS)

Dagmar KREFTING[a,b,1], Gabi ANTON[c], Irina CHAPLINSKAYA-SOBOL[a], Sabine HANSS[a], Wolfgang HOFFMANN[d], Sina M. HOPFF[e,f], Monika KRAUS[c], Roberto LORBEER[g], Bettina LORENZ-DEPIEREUX[c], Thomas ILLIG[h], Christian SCHÄFER[i], Jens SCHALLER[g], Dana STAHL[j], Heike VALENTIN[j], Peter HEUSCHMANN[k] and Janne VEHRESCHILD[e,l]

[a]*Dpt. of Medical Informatics, University Medical Center Göttingen, German Center for Cardiovascular Research (DZHK) partner site Göttingen, Germany*
[b]*Campus Institute Data Science (CIDAS), Georg-August-University Göttingen,*
[c]*Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany*
[d]*Institute for Community Medicine, University Medicine Greifswald, Germany*
[e]*Faculty of Medicine, University of Cologne, Department I of Internal Medicine, University Hospital Cologne, Germany*
[f]*Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Cologne, Germany*
[g]*Medical Heart Center and Institute of Computer-assisted Cardiovascular Medicine, Charité – Universitätsmedizin Berlin, Germany*
[h]*Hannover Unified Biobank, Hannover Medical School, Hannover, Germany*
[i]*Institute of Clinical Chemistry and Laboratory Medicine,University Medicine Greifswald, Germany*
[j]*Independent Trusted Third Party of the University Medicine Greifswald, Germany*
[k]*Institute of Clinical Epidemiology and Biometry, University of Würzburg; Clinical Trial Center, University Hospital Würzburg, Germany*
[l]*German Centre for Infection Research (DZIF), partner site Bonn- Cologne, Cologne, Department II for Internal Medicine, Hematology/Oncology, University Hospital Frankfurt, Frankfurt am Main, Germany.*

ORCID IDs: Dagmar Krefting 0000-0002-7238-5339, Gabi Anton 0000-0003-0483-892X, Irina Chaplinskaya-Sobol 0000-0003-4811-1028, Sabine Hanß 0000-0002-9133-8685, Wolfgang Hoffmann 0000-0002-6359-8797, Sina Hopff 0000-0002-8886-4596, Monika Kraus 0000-0003-4825-8724, Roberto Lorbeer 0000-0002-2224-9208, Bettina Lorenz-Depiereux 0000-0001-7215-9174, Thomas Illig 0000-0003-4284-5389, Christian Schäfer 0000-0001-8873-7231, Jens Schaller 0000-0002-9399-1653, Dana Stahl 0000-0002-4283-4543, Heike Valentin 0000-0001-6379-6534, Peter Heuschmann 0000-0002-2681-3515, Janne Vehreschild 0000-0002-5446-7170

[1] Corresponding Author: Dagmar Krefting, University Medical Center Göttingen, Germany, E-mail: dagmar.krefting@med.uni-goettingen.de

Abstract. The COVID-19 pandemic has urged the need to set up, conduct and analyze high-quality epidemiological studies within a very short time-scale to provide timely evidence on influential factors on the pandemic, e.g. COVID-19 severity and disease course. The comprehensive research infrastructure developed to run the German National Pandemic Cohort Network within the Network University Medicine is now maintained within a generic clinical epidemiology and study platform NUKLEUS. It is operated and subsequently extended to allow efficient joint planning, execution and evaluation of clinical and clinical-epidemiological studies. We aim to provide high-quality biomedical data and biospecimens and make its results widely available to the scientific community by implementing findability, accessibility, interoperability and reusability – i.e. following the FAIR guiding principles. Thus, NUKLEUS might serve as role model for FAIR and fast implementation of clinical epidemiological studies within the setting of University Medical Centers and beyond.

Keywords. clinical research infrastructure, FAIR principles, COVID-19, cohort study, Data reuse, biosamples, Images

1. Introduction

Timely availability of research results is required in the context of a pandemic, as evidence about influential factors on the severity and course of infections may have a major impact on planning efficient and effective initial treatment options and improving long-term outcomes of affected patients. But such evidence requires large-scale studies - comprehensive characterization of the patients by clinical data items, biospecimens, standardized imaging, and patients being recruited in different research settings. Typically, such large-scale studies are planned and conducted on a mid- to long-term scale, as they need significant financial resources but in particular personnel to initiate, run and analyze the study. In Germany, the *Network University Medicine* (NUM) has been founded in April 2020 to understand and master the COVID-19 pandemic by joint forces of all university medical centers. Among the 13 projects initiated in the first funding period of NUM, the National Pandemic Cohort Network (NAPKON) as the largest initiative initiated three cohort studies [1]. They differ in study population, study protocol and phenotyping details. The cohorts where complemented by overarching methodological core units. In particular during the setup phase of such multi-center studies, a fast and timely recruitment start on one hand and the fulfillment of organizational, regulatory and technical requirements for a findable, accessible, interoperable and reusable (FAIR) data collection on the other hand have been experienced as conflicting requirements. The aim of this paper is to describe the measures taken and lessons learned to reconcile these conflicts during the first funding phase and how these aspects are addressed in the NUM clinical epidemiology and study platform (NUKLEUS), established within the second funding period as a permanent collaborative research infrastructure.

2. Methods

For the successful start of the NAPKON studies, several challenges had to be addressed in a timely manner, in particular: (a) Building and supporting a large network of participating sites covering different source populations and different levels of expertise, (b) establishing a legal framework including study and regulatory documents, and

contracts between all participating infrastructure partners, and (c) coordinating implementation of the study data into the data management systems, including consent data, clinical and imaging data, and biospecimen.

We had an explicit order to make the three cohorts interoperable with each other as well as with other national and international initiatives such as the national project to share routine clinical data of COVID-19 patients [2]; and to allow for wide reuse of the collected data and biospecimen. Therefore we made specific efforts to harmonize (i) study documents - in particular patient information and informed consents, (ii) biosample collections and standard operation procedures (SOPs), (iii) common data items including quality checks and the integration of the German Corona Consensus dataset [3]; and last but not least (iv) fine-grained case fees.

The NAPKON project set up four core units for a harmonized organizational and governmental implementation of the three cohorts. The *biosample core unit* defined a core set of biospecimen to be collected by all cohorts including SOPs. The *epidemiology core unit* provides methodological consultancy of clinical epidemiological studies, delivers regular reports on data quality, and supports use and access requests and statistical analyses. The *integration core unit* evaluates external and existing cohorts upon integration into NAPKON. Last but not least, the *interaction core unit* coordinates the overall project and is in particular engaged in enabling and enhancing communication within the cohorts, the overall recruitment network, and researchers interested in data usage – reaching out to the full scientific NUM community [1].

The German Center for Cardiovascular Research (DZHK) has made available its clinical study infrastructure to NAPKON. The DZHK infrastructure encompasses integrated data management systems for informed consents, electronic case report forms, imaging data and biospecimens, respectively, with record-linkage through a trusted third party and data exports for approved requests through a transfer office. An ethics coordination supports the study coordinators in creating the study documentation and patient information suited for the intended data handling and reuse.

In 2022, we integrated the clinical study infrastructure with the *biosample*, *epidemiology* and *interaction core units* to the joint infrastructure called NUKLEUS. Within NUKLEUS, we develop and operate the research infrastructure to support joint planning, execution and evaluation of clinical and clinical epidemiological studies. The vision of NUKLEUS is to make high-quality data, biospecimens, and analysis results widely available to the scientific community. While the continuous support of the NAPKON cohorts is the main short-term objective, we consider the platform to support different multicenter clinical and clinical epidemiological studies conducted within the NUM and beyond.

3. Results

3.1. Technical infrastructure components

The technical components of the research data infrastructure encompass the following systems operated at different sites: *secutrial*® for clinical data, *CentraXX*® for biospecimen management, and *TrialComplete*® for images and biosignals. Identifying patient data and informed consents are handled by an independent trusted third party employing the open source tools *gICS*® for consent management including management

of partial or complete withdrawal, *E-PIX*® for identity management, and *gPAS*® as pseudonymization service.

The transfer office provides services to request data from the different data management systems and compiles the digital data collection. If biospecimens are part of the request, it hands over the biospecimens list to the *biosample core unit*. Biosamples are stored locally at the study sites while the metadata is centrally documented - enabling fast and coordinated access for approved usage projects. The transfer office is also responsible to compile metadata and data collections for quality assurance, reporting and accounting. Originally adopted from the DZHK, all operating sites run dedicated instances for NUKLEUS. Specific technical extensions are an export interface in *secutrial* for the harmonized data items, and a digital consent solution with *gICS*. We employ the *ProSkive* application portal for the management of data requests, with internal communication on data request handling by *GitLab* issue tracking. Further internal and external communication uses the *NAPKON-Suite*, a growing collection of collaboration and project management solutions.

3.2. Organizational and methodological infrastructure components

NUKLEUS defines and implements processes and provides methods to ensure high quality study designs, data collection and data analysis as well as efficient information and data flow within the project and with the different stakeholders. We aim at supporting both study coordinators and data users through the full research process from project definition to data reuse. For new projects, we defined an onboarding SOP and reimbursement schemes to allow a consistent calculation of the project costs. We offer methodological support to applicants for study design, biosample-related topics as well as for the statistical analysis. Informed consents are semantically annotated with a NUM-wide coding scheme to ease cross-platform data reuse. We also coordinate the comprehensive community of potentially participating sites, the so-called domain and organ specific working groups that form an independent organization body considered the specialist's backbone of NUM.

During the active recruitment phase of a study, we provide automated reporting on recruitment and data quality, as part of an accounting pipeline. Biobanking SOPs and audits at the study sites support high quality biospecimen collection.

To ease the data request process, we offer researchers who are interested in data reuse help to understand the available data and to define appropriate analysis methods. We established data request pipelines encompassing coordination of the Use & Access process, feasibility analysis, check for current consent status, data export, and finally data transfer.

3.3. Experiences from the so-far implemented studies

The three NAPKON cohorts were encouraged to start operations based on favorable reviews in August 2020 and received formal approval for funding in February 2021. The study protocols have partly been adopted from already running studies, in particular COVIDOM, LEOSS and Pa-COVID [4–6], but additional harmonization efforts and adoption of the common infrastructure were needed for all studies. First preparation steps were already ongoing since submission of the concept and funding application on July 17, 2020, and the first patient consented to the cross-sectoral cohort in Frankfurt on Nov 4, 2020, on the same day as receiving a final positive ethics vote for the study and one

day after setting the study documentation to production status. The two other cohorts recruited the first patients in the same month. All cohorts are still ongoing, with applications for data use being accepted from April 2021. To date, data from 6651 patients are available for reuse with about 32.000 reviewed visits. 83.300 primary biosamples from 5430 patients are available. Until now 104 usage applications has been approved, of which so far 64 received data. Until end of 2022, 12 sample usage applications received a total of 36.600 samples.

In April 2022, the project NU(M)KRAINE – a screening program on infectious diseases for refugees of the Ukraine was proposed. It received approval in June and started recruiting in October with almost 1800 study participants until December.

4. Discussion

The NUKLEUS study platform represents a promising infrastructure for the collection, analysis and provision of data and biosamples in the context of the ongoing COVID-19 pandemic and beyond. However, the timely implementation posed several challenges with regards to communication, workflow management, legal, ethical, data protection, and administrative challenges – in particular at study preparation. As main bottlenecks in the timely data provision for reuse we identified unclear specifications by the applicants and data not yet available. Efficient and streamlined workflow management is vital for the smooth operation of the platform, especially in a context where the integration of large amounts of data from various sources, platforms and stakeholders is a key aspect. Clear and consistent communication channels, both internally and with all stakeholders from study centers to interested researchers are essential to ensure that all participants are aware of the project's goals, objectives and their roles, as well as to facilitate the flow of information and reporting of issues in a timely manner. We are currently in discussion with further projects and will continuously explore how NUKLEUS can help collaborative research projects to be both, FAIR and FAST.

Acknowledgement: The work presented is funded by the German Ministry of Research and Education (NUM – 01KX2121).

References

[1] Schons M, Pilgram L, Reese J-P, et al. The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. *Eur J Epidemiol*. Epub ahead of print 29 July 2022. DOI: 10.1007/s10654-022-00896-z.

[2] Prokosch H-U, Bahls T, Bialke M, et al. The COVID-19 Data Exchange Platform of the German University Medicine. *Stud Health Technol Inform* 2022; 294: 674–678. DOI: 10.3233/SHTI220554

[3] Sass J, Bartschke A, Lehne M, et al. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. *BMC Med Inform Decis Mak* 2020; 20: 341. DOI: 10.1186/s12911-020-01374-w

[4] Horn A, Krist L, Lieb W, et al. Long-term health sequelae and quality of life at least 6 months after infection with SARS-CoV-2: design and rationale of the COVIDOM-study as part of the NAPKON population-based cohort platform (POP). *Infection* 2021; 49: 1277–1287.

[5] Jakob CEM, Borgmann S, Duygu F, et al. First results of the 'Lean European Open Survey on SARS-CoV-2-Infected Patients (LEOSS)'. *Infection* 2021; 49: 63–73. DOI: 10.1007/s15010-020-01499-0

[6] Kurth F, Roennefarth M, Thibeault C, et al. Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). *Infection* 2020; 48: 619–626. DOI: 10.1007/s15010-020-01464-x

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230072

Data Sharing Platform for MIMIC-IV and MIMIC-ED Data Marts: Designing a Data Retrieving System Based on the Intra-Hospital Patient Transfer Pathway

Sharareh ROSTAM NIAKAN KALHORI[a,b 1], Thomas M. DESERNO[a], Jamal SOLEIMAN[a] and Shayan KASIRI HABIBABADI[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany*
[b] *Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran*
ORCiD ID: Sharareh Rostam Niakan Kalhori https://orcid.org/0000-0002-7577-1200
ORCiD ID: Thomas Deserno https://orcid.org/0000-0003-3492-4407
ORCiD ID: Jamal Soleiman https://orcid.org/0000-0001-7842-9366
ORCiD ID: Shayan Kasiri Habibabadi https://orcid.org/0000-0003-1871-790X

Abstract. Accessibility to high-quality historical data for patients in hospitals may facilitate related predictive model development and data analysis experiments. This study provides a design for a data-sharing platform based on all possible criteria for Medical Information Mart for Intensive Care (MIMIC) IV and Emergency MIMIC-ED. Tables containing columns of medical attributions and outcomes were studied by a team of 5 experts in Medical Informatics. They completely agreed about the columns connection using subject-id, HDM-id, and stay-id as foreign keys. The tables of two marts were considered in the intra-hospital patient transfer path with various outcomes. Using the constraints, queries were generated and applied to the backend of the platform. The suggested user interface was drawn to retrieve records based on various entry criteria and present the output in the frame of a dashboard or a graph. This design is a step toward platform development that is useful for studies aimed at patient trajectory analysis, medical outcome prediction, or studies that require heterogeneous data entries.

Keywords. Platform, MIMIC, design, retrieving, prediction model, SQL

1. Introduction

Retrospectively collected medical data provide the opportunity to improve patient care through algorithm development and knowledge discovery through modeling and outcome prediction. The model has higher quality in case of being developed with proper inputs in terms of enough quantity and dimensionality [1, 2]. Thus, data sharing for model development should provide the required records based on the defined problem [2, 3].

[1] Corresponding Author: Sharareh Rostam Niakan Kalhori, Muehlenpfordtstr. 23, 38106 Braunschweig, Germany. Contact details: +49 531 391-2125, Email: sharareh.niakankalhori@plri.de.

Despite the advances in patient data collection through electronic health records (EHR), registries, and self-care apps [2], data access remains a challenge, particularly concerning big data analysis. Sharing the whole big dataset when only a part of the data is required results in adverse consequences in terms of research ethical issues and waste of time for data understanding and preparation. The well-adjusted access to medical data based on the defined study questions may overcome multifaceted concerns. Platforms are a solution to narrow the data based on the set queries and criteria [2].

The MIMIC-IV and MIMIC-ED are currently shared via the Physionet website [4] in the frame of several separate CSV files [3, 5]. They contain the data of common records with specified ID from the emergency department (MIMIC-ED) to hospital wards and ICU with clinical details (MIMIC IV) providing the possibility of following each case's transfer across the hospital and the final outcome. Currently, data for cases at hospital departments and reports for radiology, laboratory, medication, clinical notes, vital signs in chart events, history of the disease, and demographic and clinical data are provided. The MIMI-IV is the medical information for over 40,000 patients admitted to intensive care units (ICU). The newer versions of the data have been even published with more features and volume of data[3].

Although the MIMIC-III database adopted a permissive access scheme that allowed for broad reuse of the data [5], there is no data-sharing platform to manage the records according to the queries. The mechanism of schema has already been used for the prediction of key patient outcomes such as mortality, clinical deterioration, and sepsis [6]; however, the data access for a record in two different marts with the same id, to follow the patients' pathway in hospital wards or discharge, remained a lack. Furthermore, learning the metadata of the MIMIC marts and manually extracting records from the mart prolonge the accessing process. Designing a platform is a step forward to having a tool for overcoming these limitations and retrieving the data for clinical and research purposes.

Due to the entity of available data in these marts with connected records of tables via subject-id starting from ED admission to ICU discharge, and to have a logical design for the sharing data platform, the patient intra-hospital patient transfer pathway is suggested. It starts from the emergency department (ED) where the patient refers or is transferred by ambulance for further care [1, 7]. Transfer from ED to inpatient wards is a common event, with over 12 million events annually [1]. Additionally, there are four million patients are admitted to ICU each year, either from hospital wards, or ED. Most of these patients transfer from ICU to a general ward (GW) [8]. In each point of care including ED, GW, or ICU, there are possible outcomes that three of them are covered in MIMIC-ED and MIMIC-IV including transfer or admission to the next station, discharge, and death. Hence, using the patient pathway structure may facilitate patient trajectory analysis and outcome prediction by retrieving the corresponding cases. This study aimed to design a system for a data-sharing platform for MIMIC-IV and MIMIC-ED based on an intra-hospital patient transfer pathway.

2. Material and Methods

In MIMIC-ED and MIMIC IV, tables are linked by identifiers which usually have the suffix 'ID'. For example, SUBJECT-ID refers to a unique patient, HADM-ID refers to a unique admission to the hospital, and ICU stay-ID refers to a unique admission to an

intensive care unit. They are unique across the patient transfer pathway and can be used to connect the columns of the marts' tables [1]. By joining Chartevent and given outcomes such as death via ID-items, it is possible to create the constraint and have a tailored new table. To fulfill the idea of designing a SQL-based platform, a technical expert team at PLRI of TU Braunschweig in Germany was created. After one year of working with these marts for experimental purposes, the technical team started designing a platform for easier and maximum usage of the available data as a preliminary step for system development. Seven focus group meetings, every 2 hours by 5 experts in the Medical Informatics and data engineering field were conducted. After identifying the IDs as primary and foreign keys of tables' columns, the queries based on SQL were studied [9]. That is, tables based on the mart structures are defined and connected. The experts agreed on the possible constraints for the backend of the platform; based on them, the features were considered for the front end of the platform.
With complete experts' concurrence regarding the front end of the platform's design, the following functions for the platform were considered:

- Storing the data in an SQL database as it is more suitable for our use case (dynamic search),
- Using the subject-id and HDM-id as foreign keys to relate an attribute such as the vital sign in chart time while staying in a specific department with a given outcome
- Using SQL query to get the relative information based on the search criteria
- Adding the option of exporting the results to a new CSV file
- The platform should contain a graph and dashboard section to visualize the data.

Experts completely agreed on these functions to be more customized according to the patient transfer pathway in hospitals in the developing step.

3. Results

The result of the first step of checking the common data elements in MIMIC-ED and MIMIC IV is presented in table 1. These columns of the tables in two marts could be connected via the subject-id, HDM-id, and ICU stay-id of the tables. To design the backend of the platform, the information schemas were depicted. Based on table 1 and the revealed platform functions, the frontend was designed, shown in figure 2.

Table 1. The available tables of marts containing columns of medical attributions and outcomes

	Tables	Example Columns	MIMIC-ED	MIMIC-IV
Tables of Attributes	Demographic	Age, sex, Education, race	✓	
	Administration	Billing, DRG, Services	✓	
	Clinical data	Diagnosis-ICD Procedures-ICD	✓	
	Measurements	Chart events, Procedure-events, Microbiology, Lab-events, Input-output event	✓	✓
	Medications	Prescriptions medication intake	✓	
	Clinical Notes	Radiology	✓	
Outcomes	ED Admission, ED transfer, ED discharge, ED death			✓
	Ward Amission, ward transfer, ward discharge, ward death			✓
	ICU admission, ICU stay, ICU death, ICU discharged		✓	✓

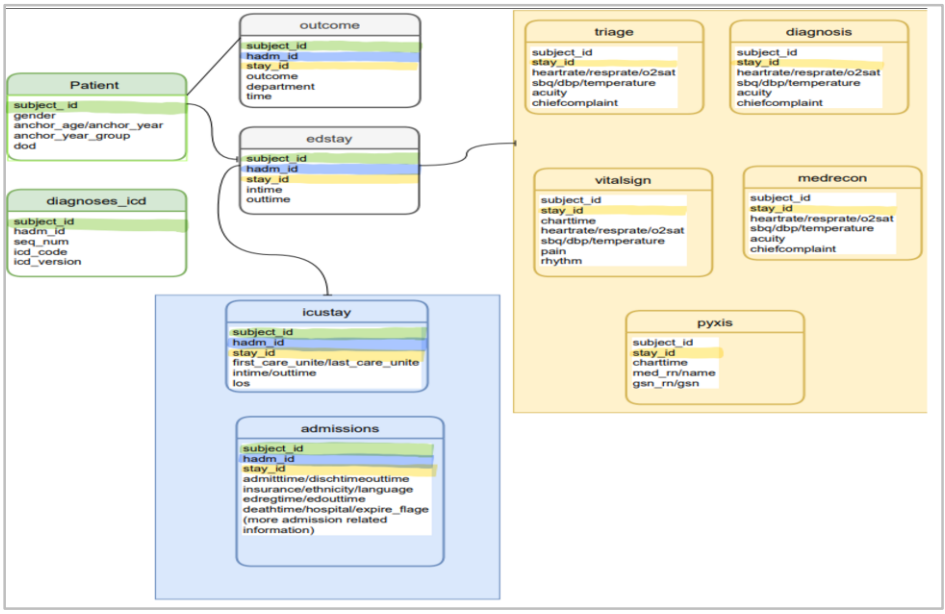


Figure 1. The example of an information schema of the backend of the platform.

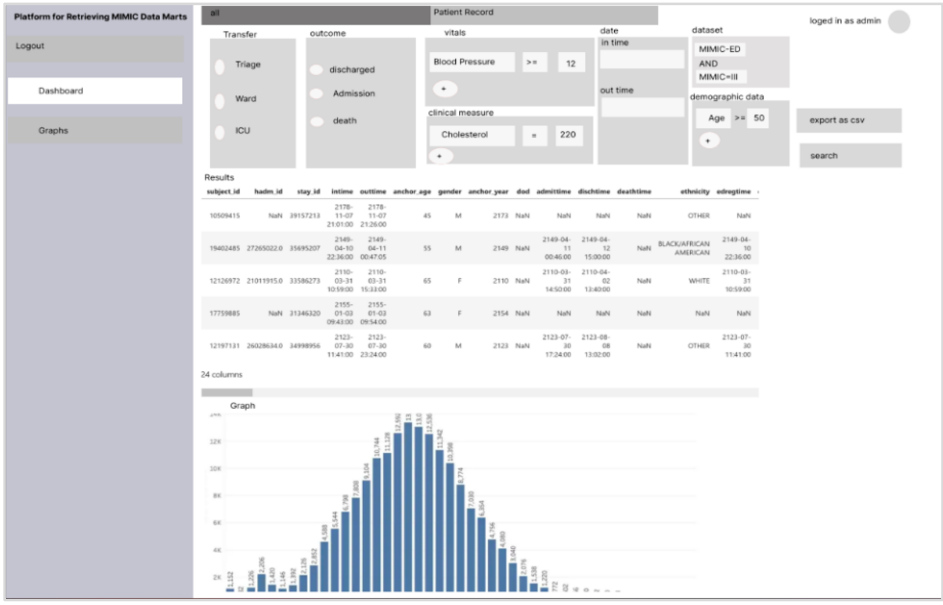


Figure 2. The user interface (UI) of the designed platform includes various data-extracting features including patient transfer, vital signs, selection of the Mart/s, date, demographic, and the given outcome. The extracted data will be provided in CSV and be shown in the frame of a graph in the dashboard section too.

4. Conclusion

According to the structure of MIMIC-IV and MIMIC-ED composed of the data of patients in the ED, triage, and admitted in the hospital (wards and ICU), designing a platform based on patient flow may be a solution for easy and quick data sharing. It is useful when the amount of data is continuously increasing. As Figure 1 shows there are several IDs presented in the same color that could connect data elements from different tables in the marts. These connections could be used for the criteria creation. As an example to get the records of a patient who died in the ICU with blood pressure greater than 140, the following steps should be done by the system:

- Getting the subject-id and stay-id of patient A to search the ICU table of MIMIC-IV
- Creating the constraints of death with BP>14 in ICU for the subject- and stay-id
- Searching the outcome column for the patients with the subject-id of patient A.
- Searching the death cases in the outcome table and picking the related record
- Using the constraints to bring up the required data for patient A in CSV.

This will be used to create all possible criteria to develop the data retrieving platform toward efficient data management and supporting researchers with heterogeneous data requirements. Examples of these studies could be titles such as prediction the ICU length of stay for patients transferred from ED with comorbidity and unstable vital signs, or emergency triage tool development to estimate the risk of ICU transfer for elderly patients affected with diabetes type II, or the trajectory prediction after hospital admission for pregnant women with unstable BP. Designing a platform as an electronic tool might be an essential need for easing the data analysis and knowledge discovery profession. However, it may face challenges regarding the limitation in data accuracy, not organized timing, and lab event data based on the unit. In the next steps, the research team has the plan for the system development and evaluation to facilitate the marts' usage.

References

[1] Xie F, Zhou J, Lee JW, Tan M, Li S, Rajnthern LS, et al. Benchmarking Predictive Risk Models for Emergency Departments with Large Public Electronic Health Records. arXiv preprint arXiv:211111017. 2021.

[2] Shortliffe EH, Chiang MF. Biomedical informatics: The science and the pragmatics. Biomedical informatics: Springer; 2021. p. 3-44.

[3] Gupta M, Gallamoza B, Cutrona N, Dhakal P, Poulain R, Beheshti R. An Extensive Data Processing Pipeline for MIMIC-IV. arXiv preprint arXiv:220413841. 2022.

[4] PhysioNet. Research Resource for Complex Physiologic Signals USA: National Institutes of Health (NIH); 1999 [cited 2023].

[5] Alistair L, Johnson E, Pollard T. Data descriptor: MIMIC-III a freely accessible critical care database. Thromb Haemost. 2016;76(2):258-62.

[6] Xia J, Pan S, Zhu M, Cai G, Yan M, Su Q, et al. A long short-term memory ensemble approach for improving the outcome prediction in intensive care unit. Computational and mathematical methods in medicine. 2019;2019.

[7] Leviner S. Patient flow within hospitals: A conceptual model. Nursing Science Quarterly. 2020;33(1):29-34.

[8] Al Owad A, Samaranayake P, Karim A, Ahsan KB. An integrated lean methodology for improving patient flow in an emergency department–case study of a Saudi Arabian hospital. Production Planning & Control. 2018;29(13):1058-81.

[9] Vathy-Fogarassy Á, Hugyák T. Uniform data access platform for SQL and NoSQL database systems. Information Systems. 2017;69:93-105.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230073

The Need for a Non-Invasive Technology for Endometriosis Detection and Care

Ariane HINE [a,1], Juliana BOWLES [a,b,2] and Thais WEBBER [a]
[a] *School of Computer Science, University of St Andrews, KY16 9SX St Andrews, UK*
[b] *SCCH, Softwarepark 32a, 4232 Hagenberg, Austria*
ORCiD ID: Ariane Hine https://orcid.org/0000-0002-7447-9371
Juliana Bowles https://orcid.org/0000-0002-5918-9114
Thais Webber https://orcid.org/0000-0002-8091-6021

Abstract. Endometriosis is a complex, poorly understood, female health condition that can markedly reduce a woman's quality of life. The gold-standard diagnostic method for Endometriosis is invasive laparoscopic surgery, which is costly, not timely, and comes with risks to the patient. We argue that the need for a non-invasive diagnosis procedure, higher quality of patient care and reduced diagnosis delay, can be fulfilled by advances and research to devise innovative computational solutions. To leverage computational and algorithmic techniques, enhanced data recording and sharing are vital. We discuss the potential benefits of using personalised computational healthcare on both the clinician and patient side, reducing the lengthy average diagnosis time (currently around 8 years).

Keywords. Female reproductive health, Endometriosis, Artificial Intelligence, Predictions models, Diagnosis time, Menstrual health

1. Introduction

Female menstrual health conditions like Endometriosis are common and, in many respects, can severely affect a woman's quality of life (QoL) [1]. The main symptoms of Endometriosis involve severe pain and fertility issues, affecting not only the reproductive organs but also frequently the bowel and bladder. For some sufferers endometriosis manifests with minor or no symptoms, which consequentially makes diagnosis using reported symptoms and clinical questionnaires even more difficult. There is a need to improve the computational detection of gynaecological conditions like Endometriosis, especially as they can be challenging to identify without invasive methods such as Laparoscopic surgery which carries physical risks and high monetary costs [2]. It is understandable that imaging modalities and physical exams are prioritised over the surgical approach due to the reduced risks and lower costs, however, this contributes to an increased diagnostic delay, disease progression, and ultimately a prolonged lower QoL for affected individuals [3]. Current research shows challenges in developing a computational-based diagnostic tool for such a complex multi-factorial condition, with non-specific symptoms that overlap with other gynaecological health conditions and different comorbidities [4]. The biggest technological barrier stems from the lack of

[1] Corresponding author: Ariane Hine, E-mail: aah4@st-andrews.ac.uk
[2] Partially supported by Austrian FWF Fellowship M 3338-N.

understanding of Endometriosis leading to the insufficiency of available, usable, and accurate data on its aetiology and pathogenesis, that could be used to construct an algorithmic solution.

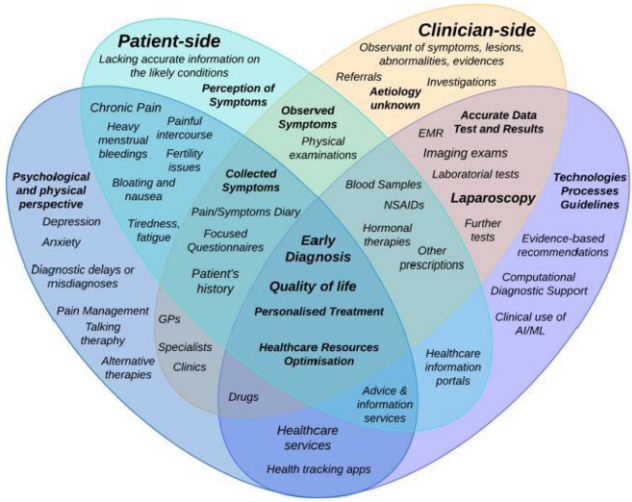


Figure 1. The context of Endometriosis diagnosis: from symptoms to improved QoL and early identification

This paper focuses on Endometriosis, one of the most poorly understood female gynaecological conditions [3]. We discuss the opportunities that the sharing of patient data can bring, regarding the diagnosis, treatment, quality of care received, and the positive implications on patient QoL. Figure 1 illustrates the context for early Endometriosis diagnosis considering the different perspectives, perceptions, and current practices on both patient and clinician sides. Section 2 describes Endometriosis as a condition, symptoms, and diagnostic delay. Section 3 presents the diagnostic techniques, implications, where technological solutions may improve the diagnosis pipeline, and where existing technologies (e.g., for recording menstrual symptoms), can improve their potential to aid in patient care. Section 4 points out how future work using data sharing can be beneficial to both clinicians and patients.

2. Endometriosis and Diagnostic Delay

Endometriosis is a common, chronic, inflammatory health condition that affects an estimated 10% of biological females of at least reproductive age. In some cases, the condition can present in an asymptomatic form, which may suggest that the rate of occurrence is greater than the predicted 1 in 10 [1]. In other cases, the condition can be life-altering, with major symptoms that have a significant potential to disrupt well-being and QoL. Clinical diagnosis (by the method of ruling out differential conditions) is being adopted more commonly in recent years due to the drawbacks of invasive surgical diagnosis, particularly the diagnostic delay incurred. On average, diagnosis takes around 8 years, with some sufferers waiting significantly longer than this [2]. Due to the progressive nature of Endometriosis, earlier diagnosis is a research priority, not only to optimise healthcare resources but also to provide better care and more personalised treatments [5]. The cause of such lengthy delays is likely due to the lack of awareness of

Endometriosis from both patient and clinician, minimal understanding of the aetiology of the condition, and typical somatisation and normalisation of its symptoms [6].

The ideal diagnostic procedure would use patient symptoms as input, but the considerably heterogeneous manifestations make this challenging, since one symptomatic presentation of Endometriosis may differ wildly from another [5]. In addition, symptoms appear to be unrelated to the disease stage, though newer research suggests that they are not necessarily unrelated to Endometriosis location and/or type [7]. The main symptoms of the condition are (but are not limited to): painful periods (dysmenorrhea), heavy bleeding during menstruation (menorrhagia), chronic/non-menstrual related pelvic pain, pain during and/or after sexual intercourse (dyspareunia), painful bowel movements (dyschezia), painful urination and loss of bladder control (dysuria), fatigue, depression or anxiety, infertility or sub-fertility, abdominal bloating and nausea, among other symptoms [1]. The variety of symptoms, and overlap with differential conditions, makes Endometriosis a challenge to identify algorithmically especially using AI techniques [8]. Despite this, we believe it is feasible, and that technological diagnosis with personalised healthcare solutions can improve the QoL for women suffering from Endometriosis.

3. Endometriosis Diagnostic Techniques & Discussion on Computational Solutions

Though clinical diagnoses are being utilised more, the gold standard diagnostic technique is still invasive surgery by means of Laparoscopy [1]. These surgeries are not only costly and risky, but their diagnostic success is highly influenced by the surgeon's experience [9]. Due to human experience and error, differences in outcome are expected, though a more uniform approach may be possible via computational input. This applies not only at the tertiary care level but also at primary care. It is also worth noting that there is a disparity between how General Practitioners (GPs) and other healthcare professionals treat menstrual symptoms, resulting in a lack of trust in the healthcare system from patients [6]. While non-invasive technologies may not be the complete solution to the long-standing issue of Endometriosis diagnosis, we believe that they provide several positive opportunities benefiting both patients and clinicians:

- Helping the general population identify irregularities in menstrual symptoms by the development of more sophisticated period tracking applications whilst reducing the diagnostic delay caused by the normalisation of gynaecological symptoms;
- Sharing of self-tracked menstrual data (with consent, in a secure manner), with medical professionals, allowing them to build a more accurate picture of a patient's symptom history. This minimises the impact of flawed human recollection of historical information that primary care doctors would receive at initial consultation;
- A clinical decision model informed by expert knowledge may assist primary care practitioners in understanding the signs and symptoms of Endometriosis and result in an earlier referral;
- Algorithmic models can be used to prioritise patients for laparoscopic surgery, considering the severity of their symptoms, QoL, and disease stage (if known). These solutions may also be used to suggest alternative solutions on an individual basis, if Laparoscopy is not appropriate for the patient;

- Prediction of endometriosis location and type in advance of Laparoscopy would be a beneficial insight for the surgical team to ensure that the correct operative experience is provisioned on a case-to-case basis to ensure the most satisfactory outcome for each patient and reduce the likelihood of the condition recurrence.

The introduction of computation into clinical care is, of course, not without intricacy. Data privacy, the sensitivity and specificity parameters used to tune algorithms, data features employed in prediction models, and acceptance of technology by clinicians are important discussion points. Any algorithmic solution can only ever be as high quality as the underlying data it makes use of. Data sharing and collaborative work within this under-researched area is crucial to successfully building innovative and accurate solutions that medical professionals will trust, accept, and incorporate into daily clinical practice.

There are several existing technologies that have the goal of improving females' menstrual well-being. In the last decade, period tracking applications have improved in functionality, and are increasingly assisting women in predicting their menstrual cycles. Though these applications aid understanding of cycles and symptoms, this self-reported data does not seem to be thorough or advanced enough for signposting to medical professionals. For example, a tracking application that notifies of menstrual irregularities generally does not have a sophisticated enough symptom-tracking interface to record the wide range of symptoms that a patient with Endometriosis may experience. Improvements in the type of data these applications collect, as well as in the sharing of this data with medical professionals, may mean that general menstrual health applications can be adapted to assist in the identification of gynaecological issues.

While research on Endometriosis has improved in the last decade, and there is a better understanding of the condition itself, we believe that there remains a lack of use of Computer Science for reducing the existing large diagnostic delay. There is a particular shortage of exploration using Artificial Intelligence (AI) including Machine Learning (ML) solutions. There is great potential to integrate ML applications into the lives of patients and/or clinicians to continue working towards a reduced time to diagnosis. Assuming this, we predict that there is an opportunity to use computation, likely an ML approach, in combination with a set of time-series patient symptoms to predict the existence of Endometriosis along with more advanced predictions such as Endometriosis type [1]. Estimating the condition stage is a more complex task due to the lack of correlation with symptoms [5]. Prediction of Endometriosis location and type in advance of Laparoscopy would be beneficial for the surgical team to ensure that the correct operative experience is provisioned on a case-to-case basis. Although having a diverse surgical team may not always be the case in every clinic, knowing that there is a high chance of a specific type of Endometriosis would be beneficial nonetheless; potentially for further referrals.

4. Conclusion

We envision that through a synergistic approach involving patient and clinician perspectives, together with technological developments, sufferers can be provided with tools and data collection interfaces to help manage their health and hence improve QoL. Not only to record a timeline of symptoms, but also to share this data so it can be examined and supplemented by clinicians for surgical triage and earlier diagnosis. The past and present recorded manifestations of the condition can be used by algorithmic

models to infer for new patients their likelihood of a gynaecological condition such as Endometriosis.

To reach this point, sharing of data and knowledge of the condition is crucial. A data collection method with notable potential is the integration of questionnaires at all levels of care, particularly at tertiary care where operative procedures are likely. The richer a picture that we can build on the symptomatic presentations (or lack thereof), and surgical characteristics (e.g., location, lesion type and size, colour), the more advanced and accurate the resulting prediction models. With enhanced quality and volume of data should come improved accuracy and trustworthiness of computational predictions; over time. Integration of these predictions into patient and clinician-side applications (e.g., clinical decision-making systems) could assist in flagging associated signs earlier, lessening the time to diagnosis. By using individual patient accounts from self-reported questionnaire data or menstrual tracking applications, together with clinician-reported medical records, cohort studies, and shared knowledge on the condition, we can expand our understanding of Endometriosis. The more data is shared over time, the more accurate and insightful will be the solutions to this poorly managed condition.

We are currently exploring ways in which we can bring data together and compare the experiences of women with Endometriosis in differing income countries and regions (within the EU and beyond), and from primary care to tertiary care. We are not dismissing the importance of clinicians and surgeons in the diagnostic and treatment pathway. We are instead suggesting that there is an opportunity to optimise the current procedures via diagnosis and care assisted by computation, providing a more personalised management plan. We envision a symbiotic solution: our ability and technology's ability to complement each other, allowing us to achieve results that neither could achieve alone. Collaborative efforts between research institutions, medical professionals, data analytics, AI and software developers will be vital in taking steps towards making this a possibility.

References

[1] The National Institute for Health and Care Excellence. NICE. Endometriosis: diagnosis and management 2017. Available from: https://www.nice.org.uk/guidance/ng73. [Accessed 13th January 2023].

[2] Pugsley Z, Ballard K. Management of endometriosis in general practice: the pathway to diagnosis. The British journal of general practice; the journal of the Royal College of General Practitioners. U.S. National Library of Medicine; 2007; 57(539):470-6.

[3] Agarwal SK, Chapron C, Giudice LC, Laufer MR, Leyland N, Missmer SA, et al. Clinical diagnosis of endometriosis: A call to action. American Journal of Obstetrics and Gynecology. 2019;220(4).

[4] Malvezzi H, Marengo EB, Podgaec S, Piccinato Cde. Endometriosis: Current challenges in modeling a multifactorial disease of unknown etiology. Journal of Translational Medicine. 2020;18(1).

[5] Rogers PA, D'Hooghe TM, Fazleabas A, Gargett CE, Giudice LC, Montgomery GW, et al. Priorities for endometriosis research: Recommendations from an international consensus workshop. Reproductive Sciences. 2009;16(4):335–46.

[6] Cox H, Henderson L, Andersen N, Cagliarini G, Ski C. Focus group study of endometriosis: Struggle, loss and the medical merry-go-round. International Journal of Nursing Practice. 2003;9(1):2–9.

[7] Lee S-Y, Koo Y-J, Lee D-H. Classification of endometriosis. Yeungnam University Journal of Medicine. 2021; 38(1):10–8.

[8] Sivajohan B, Elgendi M, Menon C, Allaire C, Yong P, Bedaiwy MA. Clinical use of artificial intelligence in endometriosis: A scoping review. npj Digital Medicine. 2022;5(1).

[9] Wood C, Kuhn R, Tsaltas J. Laparoscopic diagnosis of endometriosis. The Australian and New Zealand Journal of Obstetrics and Gynaecology. 2002; 42(3):277–81.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230074

An Annotation Workbench for Semantic Annotation of Data Collection Instruments

Julia SASSE[a] and Juliane Fluck[a,b,1]
[a] *ZB MED - Information Centre for Life Sciences* https://ror.org/0259fwx54
[b] *University of Bonn, Germany* https://ror.org/041nas322
ORCiD ID: Julia Sasse https://orcid.org/0000-0002-0660-7597,
Juliane Fluck https://orcid.org/ 0000-0003-1379-7023

Abstract. Semantic interoperability, i.e., the ability to automatically interpret the shared information in a meaningful way, is one of the most important requirements for data analysis of different sources. In the area of clinical and epidemiological studies, the target of the National Research Data Infrastructure for Personal Health Data (NFDI4Health), interoperability of data collection instruments such as case report forms (CRFs), data dictionaries and questionnaires is critical. Retrospective integration of semantic codes into study metadata at item-level is important, as ongoing or completed studies contain valuable information, which should be preserved. We present a first version of a Metadata Annotation Workbench to support annotators in dealing with a variety of complex terminologies and ontologies. User-driven development with users from the fields of nutritional epidemiology and chronic diseases ensured that the service fulfills the basic requirements for a semantic metadata annotation software for these NFDI4Health use cases. The web application can be accessed using a web browser and the source code of the software is available with an open-source MIT license.

Keywords. Interoperability, FAIR data, semantic metadata, metadata annotation

1. Introduction

The FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles for scientific data management and stewardship [1] are developed to optimize data reuse. Operationalizing this guidelines increase the use of data beyond its original purpose. For all research data collected, data descriptions and information about the corresponding variables are essential for data analysis and reuse. Plenty semi-structured and structured study documents exist in the area of clinical and epidemiological studies, that are critical for metadata collection but that are not semantically annotated. Semantic interoperability can be realized by terminology- or ontology-based semantic annotation, in which item-level metadata is enriched with terminology standards and linked to unique semantic concepts e.g. to the SNOMED Clinical Terms collection of medical terms (SNOMED CT) [2] (Table 1).

Semantic annotation is still a challenge due to various standards and semantic richness of the data. Despite of the laborious annotation process, semantic annotation is largely done manually and annotators have to manage formats of study documents and a

[1] Corresponding Author: Juliane Fluck, E-mail: fluck@zbmed.de.

variety of complex terminologies and ontologies for annotation. Therefore, a semi-automatic approach to support researchers in semantic annotation and handling the different terminologies is desirable. Several metadata enrichment tools exist [3]. Either they are domain specific (e.g. the ODMedit annotation service [4] integrated into the Medical Data Models-Portal [5]) or they are terminology specific (e.g. the SNOMED International [2] Snap2Snomed annotation service [6]). Some metadata annotation services offer only prospective metadata annotation support (e.g. the CEDAR Workbench [7]), others are not open accessible at all.

NFDI4Health [8] aims to improve the FAIR access to structured health data originating from epidemiology studies, public health and clinical studies and support the harmonization of (meta-) data. In this context, an open accessible Metadata Annotation Workbench was developed to primarily support standardized metadata annotation in the NFDI4Health use cases but with the potential to find application in other domains.

Table 1. Linking of variables to SNOMED CT [9] semantic concepts.

Subject_id	Subject_label	Object_label	Object_id	IRI
p1_3	Age	Age (qualifier value)	397669002	http://snomed.info/id/397669002
q01	How old are you?	Age (qualifier value)	397669002	http://snomed.info/id/397669002
p1_4	Gender	Gender (observable entity)	263495000	http://snomed.info/id/263495000
q02	Sex	Gender (observable entity)	263495000	http://snomed.info/id/263495000

2. Methods

The Metadata Annotation Workbench is direct accessible via a web browser [10]. The source code of the service is shared via GitHub under an open-source MIT license [11]. Next to source code, prebuild container images [12] are available.

The application is developed with the microservice architecture pattern [13]. Main rationale for the architecture of the developed service is the reuse and integration of existing terminology and ontology services that already expose semantic concepts. This separations of concerns reduces the needed development effort. Additionally, it lays grounds for future adaptions. The overall software system is depicted in Figure 1. The application consists of four services, a user interface, the aforementioned terminology service, a parsing service to read and write variable files and last the semantic annotation service. All components are interconnected using REST [14] protocol. Data is persisted in a relational database (PostgreSQL). The UI was developed in the React framework [15] with the design library Elastic UI [16] for a uniform web layout. This web application consists of a landing page with the form for uploading a file and the annotation area with the search field and a semantic concept information view.

The first core microservice, the parsing service, is responsible for the conversion of input files to an internal format for exporting the resulting annotations. The import and export file format is the column wise oriented Microsoft Excel Spreadsheet. The datasets and annotations are stored in a database within the data layer. The second microservice, the annotation service, enables the annotation of a variable with a concept of a terminology resource by associating a term or question (subject_label) with the semantic label of the concept (object_label), the concept identifier (object_id) and the International Resource Identifier (IRI) based on the Simple Standard for Sharing Ontological Mappings (SSSOM) [17] (see Table 1). For the initial search suggestion of concepts, the variable is preprocessed for querying by simple natural language processing like tokenization and stemming. The search for concepts is performed via API requests to the terminology service API of the Semantic Lookup Service (SemLookP) [18]. This service is based on software developed by the EBI [19]: the Ontology Lookup Service (OLS) [20] and the mapping service Ontology Xref Service (OxO) [21]. SemLookP serves as single point of access to the latest ontology and terminology versions of NFDI4Health and allows to annotate research data in a FAIR manner by browsing the resources through the website as well as programmatically via an API.

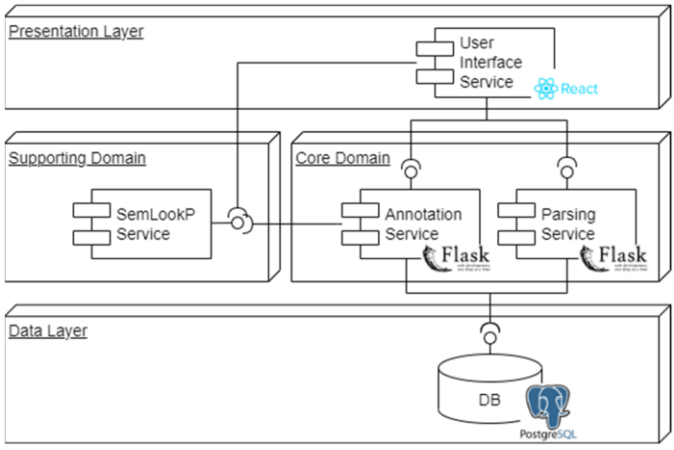


Figure 1. Metadata Annotation Workbench service interconnection.

3. Results

The web application has the following functionalities: data collection instruments can be uploaded in the convenient and common Microsoft Excel-Spreadsheet format, the user has to select a column for annotation and can select an ontology/terminology for annotation from all resources included in the terminology service SemLookP.

The metadata annotation is performed item-wise. In a first step, the user is provided initially with automatic annotation results. Detailed information about a concept is displayed in a semantic information widget that is provided by the terminology service. The user can accept the proposed annotation or modify the search term and perform a manual text search to find a matching concept. One or more concepts for the annotation of each item can be selected. Finally, the annotated instrument can then be downloaded comprising the data items and corresponding annotations.

Some terminologies such as SNOMED CT [2] or FoodEx2 [22] have terminology specific properties. SNOMED International classifies the concepts into domains indicated by a parenthetical notation at the end of a concept name. Furthermore, it supports post-coordinated expression: the required meaning is expressed by combining several concepts by logical rules. FoodEx2 uses concatenations of concepts and

additional terms (facets) describing properties and aspects of foods from various perspectives to add further detail to the information [22]. Based on user requirements of these use cases, the Metadata Annotation Workbench usability was improved. For example, annotation of a data item with multiple concepts is allowed, thereby supporting concatenation for FoodEx2 and a simple logical operation for SNOMED CT.

4. Discussion

We developed the Metadata Annotation Workbench in a user-driven development process to support standardized metadata annotation in order to foster the interoperability of data collection instruments for clinical, epidemiological and public health studies.

The terminologies for the Metadata Annotation Workbench were made available via the external terminology service SemLookP. This achieves outsourcing of the provision of the latest terminology resources and at the same time allows a replacement of the terminology service.

Several challenges for semantic metadata annotation were identified in 'nutritional epidemiology' and 'chronic diseases' use cases: the quality of the data dictionaries varies greatly and has a major influence of the automatic and human annotation. Also, the synonym content of the terminologies is important. A further challenge for ontology-based annotation is the heterogeneity of semantic annotation requirement: FoodEx2 semantic terms are often concatenations of several FoodEx2 concepts and for SNOMED CT, post-coordination might be required. The annotation service must consider terminology specifics while staying usable for various terminologies. Ontology specific functions also influences the usability for users with less expertise.

A number of necessary enhancements have been already identified:

Currently, the Metadata Annotation Workbench only supports Excel file-format as input. To enable integration of metadata annotation in workflows or data harmonization processes, other in- and output formats such as the Health Level Seven (HL7) Fast Health Interoperability Resources (FHIR) [23] format or the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM) [24] should be provided. Further future steps are the integration of advanced AI methods to optimize the pre-annotation. Also, structured usability testing will follow to enable further improvement of the service.

5. Conclusion

The Metadata Annotation Workbench fulfills the main requirements to support semantic annotation: it provides an user interface and the integration of external terminology services. Independent of the Metadata Annotation Workbench, the quality of the variable catalogs, the range of synonyms and descriptions of the terminology/ontology and their complexity are decisive factors that influence the success of semantic annotation.

Acknowledgement: We would like to thank Franziska Jannasch (German Institute of Human Nutrition (DIfE)), Carolina Schwedhelm (Max Delbrück Center for Molecular Medicine (MDC)), Carina Vorisek and Sophie Klopfenstein (Berlin Institute of Health (BIH)) for their contributions during the requirement specification.

Funding: This work was done as part of the NFDI4Health Consortium. We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 442326535.

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.

[2] SNOMED Home page. SNOMED https://www.snomed.org/ (accessed February 21, 2023).

[3] Sasse J, Darms J, Fluck J. Semantic Metadata Annotation Services in the Biomedical Domain—A Literature Review. Appl Sci 2022;12:796. https://doi.org/10.3390/app12020796.

[4] Dugas M, Meidt A, Neuhaus P, Storck M, Varghese J. ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository. BMC Med Res Methodol 2016;16:65. https://doi.org/10.1186/s12874-016-0164-9.

[5] Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P, et al. Portal of medical data models: information infrastructure for medical research and healthcare. Database J Biol Databases Curation 2016;2016:bav121. https://doi.org/10.1093/database/bav121.

[6] Snap2SNOMED https://snap.snomedtools.org/ (accessed February 21, 2023).

[7] Gonçalves RS, O'Connor MJ, Martínez-Romero M, Egyedi AL, Willrett D, Graybeal J, et al. The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments. In: d'Amato C, Fernandez M, Tamma V, Lecue F, Cudré-Mauroux P, Sequeda J, et al., editors. Semantic Web – ISWC 2017, vol. 10588, Cham: Springer International Publishing; 2017, p. 103–10. https://doi.org/10.1007/978-3-319-68204-4_10.

[8] Home. NFDI4Health https://www.nfdi4health.de/ (accessed February 21, 2023).

[9] SNOMED CT - Home https://browser.ihtsdotools.org/? (accessed March 1, 2023).

[10] Metadata Annotation Workbench https://mda.nfdi4health.de/ (accessed February 21, 2023).

[11] GitHub - nfdi4health/metadata-annotation-workbench: A metadata annotation service to support standardised metadata annotation developed by T3.2. https://github.com/nfdi4health/metadata-annotation-workbench (accessed March 1, 2023).

[12] Open Container Initiative - Open Container Initiative https://opencontainers.org/ (accessed February 21, 2023).

[13] Microservices. MartinfowlerCom https://martinfowler.com/articles/microservices.html (accessed February 21, 2023).

[14] [PDF] Architectural Styles and the Design of Network-based Software Architectures"; Doctoral dissertation | Semantic Scholar https://www.semanticscholar.org/paper/Architectural-Styles-and-the-Design-of-Software-Fielding/49fc9782483bc311c2bd1b902dfb32bcd99ff2d3 (accessed February 21, 2023).

[15] React – A JavaScript library for building user interfaces https://reactjs.org/ (accessed February 21, 2023).

[16] Elastic UI https://elastic.github.io/eui (accessed February 21, 2023).

[17] Matentzoglu N, Balhoff JP, Bello SM, Bizon C, Brush M, Callahan TJ, et al. A Simple Standard for Sharing Ontological Mappings (SSSOM). Database 2022;2022:baac035. https://doi.org/10.1093/database/baac035.

[18] Terminology Service < Semantic Lookup Platform < ZB MED https://semanticlookup.zbmed.de/ols/index (accessed February 21, 2023).

[19] Institute EB. EMBL-EBI homepage https://www.ebi.ac.uk/ (accessed February 21, 2023).

[20] Jupp S, Burdett T, Leroy C, Parkinson H. A new Ontology Lookup Service at EMBL-EBI. SWAT4LS, 2015.

[21] Jupp S, Liener T, Sarntivijai S, Vrousgou O, Burdett T, Parkinson HE. OxO - A Gravy of Ontology Mapping Extracts. In: Horridge M, Lord P, Warrender JD, editors. Proc. 8th Int. Conf. Biomed. Ontol. ICBO 2017 Newctle.--Tyne U. K. Sept. 13th - 15th 2017, vol. 2137, CEUR-WS.org; 2017.

[22] Authority (EFSA) EFS, Nikolic M, Ioannidou S. FoodEx2 maintenance 2021. EFSA Support Publ 2022;19:7220E. https://doi.org/10.2903/sp.efsa.2022.EN-7220.

[23] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst., 2013, p. 326–31. https://doi.org/10.1109/CBMS.2013.6627810.

[24] ODM-XML | CDISC https://www.cdisc.org/standards/data-exchange/odm (accessed February 21, 2023).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230075

Implementation of HL7 FHIR-Based Interoperability Profiles to Manage Care Plans for Multimorbid Patients with Mild Dementia

Mert GENCTURK[a,1], A. Emre AKPINAR[a], and Gokce B. LALECI ERTURKMEN[a]
[a] *SRDC Software Research & Development and Consultancy Corp., Ankara, Turkey*
ORCiD ID: Mert GENCTURK https://orcid.org/0000-0003-2697-5722,
Gokce B. LALECI ERTURKMEN https://orcid.org/0000-0002-6201-3849

Abstract. Management of multimorbidity in patients with mild dementia and mild cognitive impairment introduces additional challenges. The CAREPATH project provides an integrated care platform to assist both healthcare professionals and patients and their informal caregivers in the day-to-day management of care plans for this patient population. This paper introduces an HL7 FHIR-based interoperability approach for exchanging care plan action and goals with the patients and collecting feedback and adherence information from patients. In this way, seamless information exchange between healthcare professionals, patients and their informal care givers is achieved to support patients in their self-care management journey and increase their adherence to their care plans despite the burdens of mild dementia.

Keywords. Interoperability, Profiling, HL7 FHIR, Multimorbidity, Dementia

1. Introduction

Multimorbidity refers to the simultaneous presence of two or more chronic diseases in the same person. It affects more than half of the elderly population, and the number of people experiencing multimorbidity is predicted to increase by >1% per year by 2030 as the population ages [1, 2]. The aging population combined with the increasing burden of multimorbidity poses a challenge to the sustainability of healthcare systems worldwide due to high healthcare costs it causes [3]. Multimorbidity management is a complex process and involves many challenges for both the patient and their caregivers. The situation becomes more complex when multimorbidity is associated with dementia. Multimorbid patients with dementia are at additional risk compared to those without dementia, because multimorbidity can accelerate the progression of dementia and dementia can hinder effective disease management, resulting in worsening of the patient's condition [4]. Therefore, there is a need to develop solutions that addresses the challenges of multimorbidity and dementia. To address these challenges, the CAREPATH project proposes a patient-centered approach by developing a flexible and

[1] Corresponding Author: Dr. Mert Gencturk, SRDC Software Research & Development and Consultancy Corp., ODTU Teknokent Silikon Bina K1-16 06800 Cankaya, Ankara, Turkey; E-mail: mert@srdc.com.tr.

modular Integrated Care Platform consisting of several components that provide enhanced healthcare interventions for the management of multimorbidity in elderly patients living with mild cognitive impairment and mild dementia [5]. The Home and Health Monitoring Platform (HHMP) provides sensors and medical devices to continuously collect real time data of patients for monitoring their physiological and functional status. The Advanced Early Warning Smart Decision Tools (AEWSDT) processes the data gathered by the HHMP and detects declines before a clinical or functional threshold is reached. The Adaptive Integrated Care Platform (AICP) enables implementation of personalized care plans for managing multimorbidity and dementia with the help of patient's Electronic Health Records (EHRs), patient's most recent context from the HHMP, the early onset triggers delivered by the AEWSDT, and the Clinical Decision Support Services (CDSS) based on evidence based clinical guidelines. The Patient Empowerment Platform (PEP) provides personalized assistance and guidance to patients by presenting care plan activities as daily tasks to patients and collecting feedback from them for monitoring their adherence to the care plan [6]. All the data collected from the health devices, EHRs, AEWSDT, CDSS, AICP and PEP are stored in a secure Patient Data Store in HL7 FHIR format.

The Fast Healthcare Interoperability Resources (FHIR) is a health data content modelling and health data exchange standard released by HL7 to achieve interoperability with a modular approach that represents data as standalone entities called *Resources* instead of the traditional document-centric approaches [7]. It is increasingly adopted by the healthcare industry, thanks to its uncomplicated structure and easy access to resources with HTTP REST API [8]. The use of international standards such as HL7 FHIR is essential to achieve interoperability between various systems in complex architectures such as CAREPATH. However, standards alone may not be sufficient to ensure interoperability, because adaptation to specific context of use may be required depending on different system requirements. Therefore, the concept of profiling is used to add new resource elements, restrict the use of some elements, update element cardinalities, specify the terminologies to be used in specific elements, etc. [9]. Profiling can also include descriptions of choreographies, business rules and constraints, as in Integrating the Healthcare Enterprise (IHE) Profiles [10]. Although profiling is frequently used in applications in healthcare domain [11], it has also been adapted in different domains such as emergency management, disaster management and maritime surveillance [12, 13].

When it comes to multimorbidity with dementia, the care plan should focus not only on multimorbidity management, but also on tailoring it to dementia-specific needs. In addition, monitoring compliance with the care plan in dementia patients is important for managing multimorbidity effectively. In this paper, we present the implementation of HL7 FHIR-based interoperability profiles for effective management of personalized care plans in multimorbid patients living with dementia.

2. Methods

In CAREPATH, care plan management of multimorbid patients with dementia is mainly handled by two components: AICP and PEP. AICP is a web-based platform that allows creation and sharing of personalized care plans across multi-disciplinary care teams, including health and social care providers. In the management of multimorbidity, it is of utmost importance for patients to take health measurements at home, take their medications as prescribed on time, follow lifestyle recommendations, and have regular

follow-up visits. In this regard, a personalized care plan is created by the care team of the patient based on patient's available health records such as diseases, symptoms, drugs, vital signs, recent lab results, family history etc. in the FHIR repository, which was already retrieved from underlying EHR systems or generated by the CAREPATH components. In the care plan, each item that patient needs to perform is defined as an "activity". In addition, targets such as weight or blood pressure can also be defined and added to the care plan as "goals".

The medical data of patients are accessed via the *Condition, Observation, MedicationStatement* and *FamilyHistory* FHIR resources. Care plan of a patient is stored in the *CarePlan* resource. In *CarePlan*, the goals are modelled with *Goal* resource, while activities are defined in *ServiceRequest*, *MedicationRequest*, and *Appointment* resources. *ServiceRequest* is used for self-measurements of vital signs and lifestyle recommendations. The medications that the patient will start/stop using or the dosage changes in already used medications are specified in *MedicationRequest*. *Appointment* is used for setting a follow-up appointment. However, considering the fact that patients living with dementia often battle memory problems and have difficulty in performing daily activities due to cognitive decline, it would not be sufficient to list the activities to be done in the care plan alone. Therefore, several interoperability profiles are defined with *ServiceRequest* resource, in which the exact time or period in a day should be specified in one of the "occurencePeriod" or "occurenceTiming" elements. In addition, these patients need to take a lot of medications during the day due to their multimorbidity, but it is not practical to list these one by one in the care plan. Instead, it is more practical to divide the day into parts and ask patients to take their medication with the pill box method, e.g., "take your medication before breakfast". Therefore, in the *CarePlan* profile, the *MedicationRequest*s are encapsulated in *ServiceRequest* objects and shown in the care plan as medication intake activity. Furthermore, to aid patients with dementia for remembering daily activities outside the care plan such as meeting with a friend, a custom task interoperability profile is defined with *ServiceRequest* resource.

In the management of multimorbidity in patients living with dementia, it is also important to follow patient's adherence to the care plan activities. In this regard, a mobile application called PEP is provided both to the patients and their informal caregivers. In PEP, patients display the daily activities they need to perform in a day with time and other useful information that can help them to perform corresponding activity. Then, either the patients or their informal caregivers provide feedback on whether the activities are performed or not. These feedbacks are stored in an *Observation* FHIR resource based on the activity feedback profile. The CAREPATH interoperability profiles are presented and explained in detail in the next section.

3. Results

Figure 1 shows the overview of the FHIR resources and references for CAREPATH interoperability profiles. In the figure, only the extensions to the original FHIR resources, cardinality updates and restrictions on the terminology systems and values are presented. Other elements in a resource can still be used as defined in its FHIR documentation.

As shown in the figure, at the heart of the CAREPATH interoperability profiles is the *CarePlan* profile. A care plan is created by a team of caregivers for a specific patient. For these, *CareTeam* and *Patient* FHIR resources are used without any modification. The "intent" element in the *CarePlan* is always set as "plan" and goals are stored in *Goal*

resource as it is defined in FHIR. For *Conditions* and *MedicationStatements*, ICD-10 and ATC international coding systems are used, respectively. In *Observation*, LOINC is used for laboratory results, while SNOMED-CT is used for lifestyle parameters.

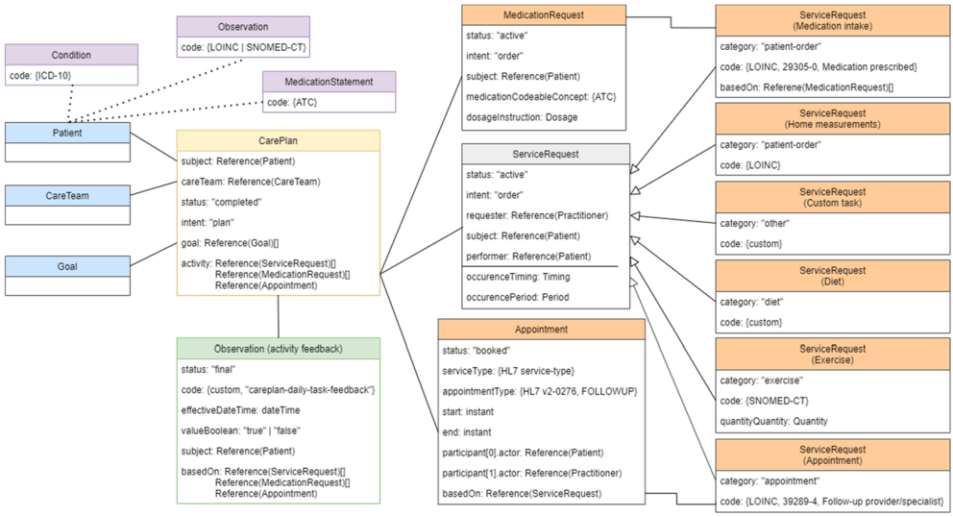


Figure 1. Overview of the FHIR resources and references for CAREPATH interoperability profiles.

For the *ServiceRequest* activities in the care plan, 6 profiles are defined. These profiles extend the base *ServiceRequest* profile shown in grey in the figure. The most important customization in the base profile is that at least one of the "occurenceTiming" and "occurencePeriod" elements are mandatory. If the "occurenceTiming" is used, then the "frequency", "period", "periodUnit", and "when" elements in it must be provided. The medication intake *ServiceRequest* profile is used for indicating when a box of medication is to be taken in a day as described before. The LOINC code of "29305-0" is used and the list of medications are referenced in "code" and "basedOn" elements, respectively. The details of medications are provided in *MedicationRequest* with ATC code and dosage information as defined in *MedicationRequest* profile. The home measurement *ServiceRequest* profile requires the use of a LOINC code in the "code" field (such as "29463-7" for body weight), whereas the custom task and diet *ServiceRequest* profiles requires the use of some custom codes, because there is no standard code system available for custom tasks or dietary recommendations. In the exercise *ServiceRequest* profile, SNOMED-CT code system is used for the "code" element, and information like duration and repetition are provided in "quantityQuantity". In the care plan, follow-up appointments are stored in *Appointment* FHIR resource as shown in the figure, but to show these appointments in the daily activities of patients and collect feedback about them, an appointment *ServiceRequest* profile is defined and linked to corresponding *Appointment* as in the case of medications. In this profile, the LOINC code of "39289-4" is used to indicate follow-up appointment. The types of activities in *ServiceRequest* are specific in the "category" element, such as "patient-order", "diet" and "exercise". Finally, for the feedbacks of the patients and/or their informal caregivers on performance of daily activities, activity feedback *Observation* profile is defined, where the "valueBoolean" is used to indicate whether task is performed and "basedOn" is used to refer corresponding activity.

4. Discussion and Conclusion

In this paper, the implementation of HL7 FHIR-based interoperability profiles for effective management of personalized care plans in multimorbid patients living with mild dementia is presented. We chose to base our interoperability architecture on the HL7 FHIR, a widely accepted international standard in healthcare domain, to increase its future exploitation in integration with existing health IT systems at care sites. We will conduct a Technical Validation and Usability (TVU) study and Clinical Investigation (CI) at four different pilot sites in Europe to demonstrate the effectiveness of the CAREPATH interoperability profiles presented in this paper in real life care settings.

Acknowledgements

The work presented in this paper have received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No 945169.

References

[1] Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, Meinow B, Fratiglioni L. Aging with multimorbidity: a systematic review of the literature. Ageing research reviews. 2011 Sep 1;10(4):430-9, doi: 10.1016/j.arr.2011.03.003

[2] Yoon J, Zulman D, Scott JY, Maciejewski ML. Costs associated with multimorbidity among VA patients. Medical care. 2014 Mar;52(Suppl 3):S31, doi: 10.1097/MLR.0000000000000061

[3] Ofori-Asenso R, Chin KL, Curtis AJ, Zomer E, Zoungas S, Liew D. Recent patterns of multimorbidity among older adults in high-income countries. Population health management. 2019 Apr 1;22(2):127-37, doi: 10.1089/pop.2018.0069

[4] Calderón‐Larrañaga A, Vetrano DL, Ferrucci L, Mercer SW, Marengoni A, Onder G, Eriksdotter M, Fratiglioni L. Multimorbidity and functional impairment‐bidirectional interplay, synergistic effects and common pathways. Journal of internal medicine. 2019 Mar;285(3):255-71, doi: 10.1111/joim.12843

[5] CAREPATH: Developing Digital Integrated Care Solutions for Multimorbid Patients with Dementia. In: Advances in Informatics, Management and Technology in Healthcare 2022 (pp. 487-490). IOS Press, doi: 10.3233/SHTI220771

[6] Gencturk M, Laleci Erturkmen GB, Gappa H, Schmidt-Barzynski W, Steinhoff A, et al. The design of a mobile platform providing personalized assistance to older multimorbid patients with mild dementia or mild cognitive impairment (MCI). In: 10th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2022); 2022 Sep 2; Lisbon, Portugal. ACM, New York, NY, USA, doi: 10.1145/3563137.3563138

[7] Using HL7 FHIR to achieve interoperability in patient health record. Journal of biomedical informatics. 2019 Jun 1;94:103188, doi: 10.1016/j.jbi.2019.103188

[8] Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. JMIR medical informatics. 2021 Jul 30;9(7):e21929, doi: 10.2196/21929

[9] FHIR Profiling, https://hl7.org/fhir/profiling.html

[10] Integrating the Healthcare Enterprise (IHE) Profiles, https://www.ihe.net/resources/profiles/

[11] Erturkmen GB, Yuksel M, Baskaya M, Sarigul B, Teoman A, Yilmaz G, De Manuel E. Interoperability Architecture of the ADLIFE Patient Empowerment Platform. Public Health and Informatics 2021 (pp. 936 - 941). IOS Press, doi: 10.3233/SHTI210316

[12] Gencturk M, Arisi R, Toscano L, Kabak Y, Di Ciano M, Palmitessa A. Profiling Approach for the Interoperability of Command & Control Systems with Sensing Systems in Emergency Management. In: IWEI Workshops 2015.

[13] Gencturk M, Duro R, Kabak Y, Božic B, Kahveci K, Yilmaz B. Interoperability profiles for disaster management and maritime surveillance. In: eChallenges e-2015 Conference; 2015 Nov 25; Vilnius, Lithuania. IEEE. p. 1-9, doi: 10.1109/eCHALLENGES.2015.7441073

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230076

Towards Prediction of Injuries in Traffic Accidents

Viktor M. G. SOBOTTA[a,1], Maximilian FLORMANN[b,], Roman HENZE[b] and Thomas M. DESERNO[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, 38106 Braunschweig, Germany*
[b] *Institute of Automotive Engineering, TU Braunschweig, 38106 Braunschweig, Germany*

Abstract. For people involved in road traffic accidents, the time necessary to respond is crucial and it is hard to discern, which persons in which cars most urgently need help. To plan the rescue operation before arriving at the scene, digital information regarding the severity of the accident is vital. Our framework aims to transmit available data from the in-car sensors and to simulate the forces enacted on occupants using injury models. To avoid data security and privacy issues, we install low-cost hardware in the car for aggregation and preprocessing. Our framework can be retrofitted to existing cars and therefore could extend the benefits to a wide range of people

Keywords. Injury prediction, Traffic accident simulation, Framework

1. Introduction

Road traffic crashes are one of the leading causes of death for children and young adults. The World Health Organization (WHO) estimates 1.3 million fatalities globally each year [1]. Sanchez-Mangas et al. have shown that a fast response to traffic accidents is vital to increase the chances of survival [2]. Accordingly, governments regulate the methods of emergency calls. Modern approaches such as eCall and Next Generation 112 focus on the automatic response to crashes, removing the necessity for the error-prone intervention by a person or the time-consuming dialogue with the dispatcher [3,4]. Such protocols also include further data to support the planning and execution of the emergency response mission. Since traffic accidents often involve multiple vehicles [5], it is furthermore crucial to provide information to the emergency responding team, which assists them in the reconstruction of the accident and prediction of the resulting injuries. This would enable them to provide the fastest aid to the persons most in need and therefore would lead to the best outcome.

Instead of relying on custom solutions, which would take time to permeate the markets and also exclude the vast amount of existing cars, we utilize existing sensors build-in cars and their recording and communicating facilities. The Controller Area Network (CAN) is a wired vehicle bus architecture. Since 1986, it is the standard in automotive

[1] Corresponding Author: Viktor Maria Georg Sobotta, E-mail: viktor.sobotta@plri.de

engineering. The CAN bus interconnects the various sensors and actors within the vehicle and its communication protocol provides error correction. Modern cars feature a variety of sensors, which as identified by Deserno et al. could be crucial for the prediction of injuries [6]. Through the extraction of this data via an interface and subsequent semantic mapping, crash-related information could be obtained for existing cars indifferent of model and manufacturer.

Data broadcasting can be integrated into the International Standard Accident Number (ISAN), which provides the semantically interoperable information exchange between automatic systems of the emergency alerting, responding, and curing instances [7]. Tri et al. used the BeamNG physics engine to virtually reconstruct road accidents [8]. The observed forces acting on the simulated passengers can be applied to human injury models, to produce a comprehensive representation of the possible state of each person involved [9]. In this paper, we suggest a framework for information merging, visualizing, and presenting to aid the rescue personnel's decision-making process prior to the arrival of the rescue team at the accident site.

2. Methods

2.1. Data extraction

Our method depends on the reliable extraction of relevant data from the CAN-bus. This requires the design of an interface for data processing and semantic mapping. Sudarshan et al. [10] utilize the Raspberry Pi and PiCan2 to interface the CAN bus of vehicles. The Raspberry Pi can perform simple computing and interpretation of the provided CAN-bus data in real-time and functions as an extendable hub. To extract the data model and manufacture-independently, we use a look-up table for each car model locally on the Raspberry Pi. The mapping of different sensors and sampling rates to the specified data fields is required for the subsequent simulation. To capture all relevant data, the Raspberry Pi continuously reads information from the CAN bus in a certain time frame. The reporting of a crash (alerting) is handled by the eCall system or, if no eCall exists in a vehicle, the alerting is triggered by the deployment of the airbags. Once our system detects an event, the Raspberry Pi gathers the data frame prior to and all frames during the event. The Raspberry Pi then establishes a connection with the ISAN network, encodes the data for transmission, and submits the data payload.

2.2. Data transmission

Figure 1 illustrates the communication of the individual components of our framework. First, the crash is registered and then the encoded crash data is sent via the ISAN protocol. The data is received by the ISAN system handling verification, encryption, and decryption of data. In the following step, the data is relayed to the simulation server that executes the physical simulation and applies the injury model. The so generated data then gets visualized on the server and can be accessed via an authentication process. The simulated data for the occurred accident is referenced by the data linking process provided in the ISAN specifications [7].

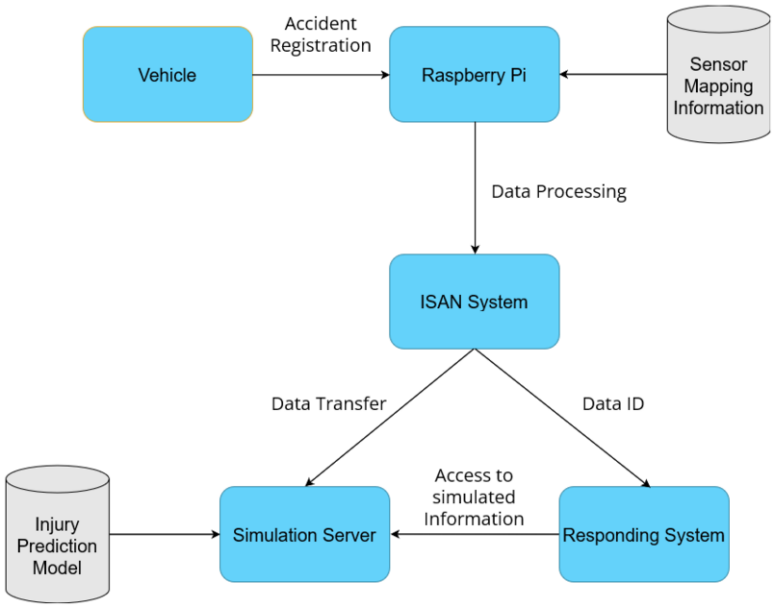


Figure 1. Overview of the proposed framework.

2.3. Injury prediction and visualization

To gauge the extent of the injuries inflicted, we use the received data to simulate the physical forces acting on the passengers of the involved vehicles. To implement this, the received data on the simulation server creates a simulation instance for the recorded accident, linking the incoming data of each vehicle by spatial and temporal proximity and assigning a unique identifier of the crash to each car and passenger. For each car within the instance, we run a simulation based on the reported data using a yet-to-be-specified physics engine. The vehicle model used in the simulation is a stand-in model based on the reported vehicle class. If the number of occupants is not provided, we assume full occupancy for the simulation and highlight non-verified passengers with flags accordingly. To execute the physical simulation, we apply the registered forces to the stand-in vehicle in which we place a virtual ragdoll dummy on all occupied seats to capture the impact and forces acting on the passengers. For the design of the ragdoll multiple variants, accommodating the different sexes and statures can be used. Depending on the simulation load of the server, multiple instances of the simulation can be run to generate data for the different dummies. Following the simulation of the experienced forces, a human injury model will be utilized to translate the applied forces into the resulting injuries.

3. Results

The prior steps culminate in a system, which provides the information for the identification of the persons most affected by the accident. To make the data accessible we suggest the following visualization.

Figure 2. Left: Real-time render of BeamNG.drive with dummy placed on driver seat **Right:** Visualization of a possible representation of forces enacted on dummy during crash.

First, provide an overview of the entire scene of the accident with all cars involved. This will be implemented, by using the stand-in vehicles and the transmitted GPS location. A tier system, assigning colors to the severity of the simulated injuries will be employed to visualize the vehicles with passengers in critical condition. Selecting a vehicle then provides a view of all occupants represented by the corresponding 3D dummy. As shown in Figure 2, a heat map of the experienced forces can be assigned as a texture, and selecting a dummy produces the list of assumed injuries derived from the injury model.

4. Discussion

The here proposed framework for the prediction of road traffic accidents could provide a scale-able and cost-efficient approach to visualize existing sensor information collected even by older vehicles. The use of a Raspberry Pi for the local aggregation and semantic encoding allows for an undemanding extension with further sensors in the future. With this, we achieved our set goal for the accessibility of our framework, which could be particularly interesting for developing countries, which exhibit heightened numbers of car accidents [1]. Our suggested framework builds on the previously established research, that demonstrated the validity of the reconstruction of accidents using a state-of-the-art physics engines [8] and the accurate prediction of injuries through human finite element models (FEMs) [9,11]. While information not recorded by the sensors prohibits a perfect reconstruction of the accident, the integration of auxiliary technologies, like wearables could alleviate the issue. We expect the system to provide information to emergency response teams, from previously unused sensors, which can function as a lower bound for injury prediction and could reduce the mortality of persons involved in road traffic accidents. Going forward, future works need to implement and validate the herein-proposed components

of the framework. This includes the design of the real-time semantic processing, encoding, and transmission of the data locally in the car. The incorporation of the communication process within the existing ISAN framework. The implementation of the simulation server, encompassing the inclusion of an appropriate physics engine that allows for the extraction of the relevant data. Lastly, all produced information needs to be accessible in an intelligible format, which allows for a hierarchical overview of the whole scene, starting with vehicles and ending with the detailed injuries afflicted on a single passenger.

Following the creation of a prototype, the practicalities of the framework can be further examined. Hardware requirements can be specified and scale-ability regarding the number and complexity of simulations could be evaluated. Existing databases, such as the Crash Injury Research and Engineering Network (CIREN) which have been employed by Golman et al. to test the injury prediction of FEMs[11], could be utilized to test the injury prediction process. Expanding on this approach, the proposed visualization of information could be scrutinized. Finally, it needs to be tested, whether the provided information leads to an improvement in the decision making process and has a measurable impact on the outcome of rescue missions.

5. Acknowledgment

This research has partly been funded by Amt für regionale Landesentwicklung (ArL) Leine-Weser (D2-06023-053-008/DiRekt).

References

[1] World Health Organization. Global status report on road safety 2018. Geneva, 2020.
[2] Sanchez-Mangas R, Garcıa-Ferrrer A, de Juan A, Arroyo AM. The probability of death in road traffic accidents: how important is a quick medical response? Accident Anal Prevent. 2010; 42(4): 1048-56.
[3] Bonyar A et al. A review on current eCall systems for autonomous car accident detection. Proc IEEE ´ ISSE 2017. pp. 1-8 .
[4] Sedlar U et al. Next generation emergency services based on the pan-European mobile emergency application (PEMEA) protocol: leveraging mobile positioning and context information. Wireless Comm Mobile Comp. 2019;
[5] Hiroshi H, Fischbeck PS, Fischhoff B. Traffic accident statistics and risk perceptions in Japan and the United States. Accident Anal Prevent. 2000; 32(6): 827-35.
[6] Deserno T M, Stachowiak N, Wang J. Beyond a Three-Dimensional Digital Rescue Sheet. 2021 (ICEC-CME) (pp. 1-6). IEEE
[7] Spicher N et al. Proposing an international standard accident number for interconnecting information and communication technology systems of the rescue chain. Methods Inform Med. 2021; 60(S1): e20-e31.
[8] Tri H, Gambi A, Fraser G. AC3R: automatically reconstructing car crashes from police reports. Proc IEEE ICSE. 2019. pp. 31-34.
[9] Chao Y et al. A computational biomechanics human body model coupling finite element and multibody segments for assessment of head/brain injuries in car-to-pedestrian collisions. Int J Environment Res Public Health. 2020; 17(2): 492.
[10] Sudarshan P, Sangdon L. Design and implementation of a CAN data analysis test bench based on Raspberry Pi. J Multimed Inf Syst 2019;6(4):239-244.
[11] Golman, Adam J., et al.: Injury prediction in a side impact crash using human body model simulation. Accident Analysis & Prevention 64 (2014): pp. 1-8.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230077

Predicting Family Implementation of Complementary and Alternative Medicine in Autism Online Communities

Zhenni Ni[a, b,1] , Yuxing Qian[a], Hao Li[c] and Zhizhen Yao[a]
[a] *School of Information management, Wuhan University, China*
[b] *LIMICS (UMR_S 1142), Sorbonne Université, France*
[c] *Sch Med & Hlth Management, Huazhong Univ Sci & Technol, China*

Abstract. Complementary and alternative medicine (CAM) is widely adopted by families with autistic children. This study aims to predict family caregivers' CAM implementation in Autism online communities. Dietary interventions were reported as a case study. We extracted behavioral (degree and betweenness), environmental (positive feedback and social persuasion), and personal features (language style) of family caregivers in online communities. The results of the experiment showed that random forests performed well in predicting families' tendency to implement CAM (AUC=0.887). It is promising to use machine learning to predict and intervene in the CAM implementation by family caregivers.

Keywords. Autism spectrum disorder, Machine learning, Adoption behavior

1. Introduction

The use of complementary and alternative medicine (CAM) for children with Autism spectrum disorder (ASD) is common around the world [1]. Families are particularly susceptible to poorly evaluated fads and exaggerated claims about the effectiveness of so-called alternative treatments in social media [2]. Previous studies have revealed the prevalence, reasons and risk of CAM use in children with ASD [3]. Utilizing social media data to predict the implementation of CAM by families can help take steps to monitor the spread of these treatments. In this study, we report a case study of dietary interventions. Machine learning models are applied to predict CAM implementation of families with autistic children in Online Communities.

Table 1. Features and measurements for the prediction of family CAM implementation

Dimensions	Features	Measures
Behavior	Social interaction	Degree centrality [a]
		Betweenness centrality [a]
Environmental factors	Positive feedback	Number of posts that provide positive feedback
	Social persuasion	Number of replies from opinion leaders [b]
Personal factors	Language style	Percentage of words in TextMind [c]

[a] A weighted directed network was constructed based on the user's online interaction behavior.
[b] We calculated the PageRank score for each user. The top 1% of users were identified as opinion leaders.
[c] TextMind is a text analysis tool that can be used to reveal different emotions, thinking styles, social concerns, and parts of speech.

[1] Corresponding Author: Zhenni Ni, E-mail: Jennie_n@whu.edu.cn.

2. Method and Results

Our dataset covered 136,763 records (60,479 posts and 76,284 comments) in the biggest Chinese autism community from January 2017 to May 2019. A preliminary selection was performed with keywords related to dietary interventions (10,583 records were obtained). Following that, we randomly chose 3000 records and manually labeled whether they were related to dietary interventions. The remaining records were automatically classified with BERT. Users were identified as family caregivers if any of their posts contained phrases such as "my daughter". Finally, we got 3279 records related to dietary interventions posted by unique 982 family caregivers. Inspired by previous studies [4], we extracted some features from three dimensions: behavior, environmental factors, and personal factors to predict family implementation (Table 1). We manually reviewed the posts of each family caregiver. Family caregivers who self-reported implementation or provided feedback were marked as implemented; others were marked as not implemented. FeatureSelector package in Python selected meaningful features from 102 features in TextMind. We applied four machine learning algorithms (Table 2), including Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). AUC and Loss are chosen as evaluation metrics because they indicate how well the model can distinguish between classes. Random Forest has the best performance.

Table 2. Performance of algorithms

SVM		**RF**		**LightGBM**		**XGBoost**	
AUC	Loss	AUC	Loss	AUC	Loss	AUC	Loss
0.827	6.193	**0.887**	4.287	0.870	6.193	0.868	6.193

3. Discussion and Conclusions

The findings reveal that machine learning can predict the CAM implementation of family caregivers based on their online characteristics. Our study makes it possible to provide targeted online decision support to families implementing CAM. This could help families make informed decisions about the health of their children. For example, online interventions could be developed to improve access to evidence-based practices, increase caregiver education about evidence-based practices, and reduce the use of CAM that may be ineffective or harmful. In the future, further research will be necessary to eliminate possible biases. The designed therapies need to be in consensus in the expert domain and then recommended to the caregivers online.

References

[1] Zhukova MA, et al. Complementary and Alternative Treatments for Autism Spectrum Disorders: A Review for Parents and Clinicians. Clinical Psychology and Special Education. 2020;9(3):142-73.
[2] Lord C, Charman T, Havdahl A, Carbone P, Anagnostou E, Boyd B, et al. The Lancet Commission on the future of care and clinical research in autism. The Lancet. 2022;399(10321):271-334.
[3] Trudeau MS, Madden RE, Parnell JA, Ben Gibbard W, Shearer J. Dietary and Supplement-Based Complementary and Alternative Medicine Use in Pediatric Autism Spectrum Disorder. Nutrients. 2019;11(8).
[4] Ni ZN, Qian YX, Yao ZZ, Zhang S. Understanding the Adoption of Dietary Interventions Within a Chinese Autism Online Community: A Diffusion of Innovations Perspective. Health Commun. 2022.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230078

Integrating Cross-Departmental Data

Julia GEHRMANN [a,1] and Oya BEYAN [a]
[a] *Institute for Biomedical Informatics Cologne, Germany*
ORCiD ID: Julia Gehrmann https://orcid.org/0000-0002-4101-5458,
Oya Beyan https://orcid.org/0000-0001-7611-3501

Abstract. Developing smart clinical decision support systems requires integrating data from several medical departments. This short paper outlines the challenges we faced in cross-departmental data integration for an oncological use case. Most severely, they have led to a *significant reduction in case numbers*. Only 2,77% of cases meeting the initial inclusion criteria of the use case were present in all accessed data sources.

Keywords. Medical Data, Oncology, Integration, Challenges, Germany.

1. Introduction

Smart clinical decision support systems (CDS) promise to improve future health care [1,2]. However, they can only make well-founded decisions when provided with sufficiently rich training data [3]. Aiming for a prototypical CDS we started our collaboration with the Center for Integrated Oncology (CIO) Cologne. As a first step we compiled a cross-departmental dataset containing *routine clinical data* of oncological cases comprehensive enough to reliably identify the respective type of cancer. The CIO's own *Clinical Cancer Registry* (CCR), the *radiology department* of University Hospital Cologne (UHC) and the *Medical Data Integration Center* (MeDIC) Cologne served as data providers.

2. Methods

Starting from the CCR export (diagnoses and histological examination results from 2000 to 2021), we filtered the diagnoses based on the use case requirements and forwarded the Patient Identifiers (PIDs) linked to the remaining diagnoses to the other data providers. They exported available data for these PIDs from the *Radiological Study System* (RSS) and the *Laboratory Information System* (LIS). The individual exports were preprocessed by calculating metafeatures and integrated based on the PIDs. In case of several eligible examination results for a single case we chose the result determined closest in time to the diagnosis. Complying to the requirements of the data protection office (DPO) and the ethics committee (EC), all preprocessing steps were implemented by us in python scripts, but executed by the oncologists of CIO Cologne.

[1] Corresponding Author: Julia Gehrmann, E-mail: julia.gehrmann1@uk-koeln.de.

3. Results

During data integration we observed a *significant decrease in case numbers*. While the original CCR export contained 14144 eligible diagnoses, only 392 of them were present in all exports (histology export: 139878, LIS export: 12105, RSS export: 410). Additionally, we noticed *high heterogeneity* in the histology and LIS exports manifesting itself both in *high missingness rates* for some features and several *examination results* for single cases. This challenged the linkage of the individual exports just as the *inconsistent IDs*. While the CCR uses own diagnosis IDs, the lab export links the results to an administrative case ID. To still enable linkage of the exports we based it on the PIDs and the dates.

4. Discussion

The largest decrease in case numbers was incurred by using the RSS as a data source. It only contains data of *study participants*. On the other hand, it is the only source providing tabular radiological data at UHC. In future it might be worth applying advanced techniques like image recognition and natural language processing to obtain structured data while keeping case numbers high. Inherent specificities of medical real-world data like *inconsistencies, heterogeneity and sensitivity* cannot be removed systematically but the experience we gained in this use case will support us in dealing with them in future. Moreover, we got to know the structure of the system exports which will ease the *blind data integration*, i.e. writing preprocessing scripts executed by physicians, in upcoming projects.

5. Conclusion

Integrating cross-departmental medical data includes a variety of challenges. These result from necessary *data protection, the primary-care focus of data systems* and *specificities of real-world medical data* like heterogeneity and sparsity. Despite these challenges, we successfully compiled a cross-departmental dataset which can now be used to build a prototype for a CDS in oncology. For further projects aiming for cross-departmental data integration, we recommend starting with as few exclusion criteria as possible because the integration process itself will most likely lead to a significant decrease in case numbers.

References

[1] Gruson D, Helleputte T, Rousseau P, Gruson D. Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. Clinical biochemistry. 2019;69:1-7. doi: https://doi.org/10.1016/j.clinbiochem.2019.04.013

[2] Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC medicine. 2018;16(1):1-15. doi: https://doi.org/10.1186/s12916-018-1122-7

[3] Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nature Reviews Cancer. 2022;22(2):114-26. doi: https://doi.org/10.1038/s41568-021-00408-3

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230079

Classification of Parkinson's Disease from Voice - Analysis of Data Selection Bias

Alexander BRENNER[a,1], Catharina Marie VAN ALEN[a], Lucas PLAGWITZ[a] and Julian VARGHESE[a]
[a]*Institute of Medical Informatics, University of Münster, Münster, Germany*

Abstract. A growing number of studies have been researching biomarkers of Parkinson's disease (PD) using mobile technology. Many have shown high accuracy in PD classification using machine learning (ML) and voice records from the mPower study, a large database of PD patients and healthy controls. Since the dataset has unbalanced class, gender and age distribution, it is important to consider appropriate sampling when assessing classification scores. We analyse biases, such as identity confounding and implicit learning of non-disease-specific characteristics and present a sampling strategy to highlight and prevent these problems.

Keywords. Machine Learning, Parkinson's Disease, Selection Bias

1. Introduction

The incidence and global burden of Parkinson's disease (PD) is increasing [1]. PD is associated with a variety of symptoms, including tremors, rigidity, changes in speech and gait, and non-motor symptoms. Early treatment can reduce burden, but screening for PD can be time-consuming. The mPower study provides a large dataset of PD patients and healthy controls (HC) including voice records from smartphones [3]. Based on this data various studies reported accuracies of up to 90% in the distinction of PD from HC using machine learning (ML) [1,2]. Still, a common problem are repetitive samples of the same individual. Many analyses lack declaration of subject-wise train/test splits, which may lead to identity confounding. In addition, controls should represent clinical practice. Otherwise, models may differentiate people by age instead of disease-specific patterns. We present an approach to identify bias and to derive fair scores with stratified sampling.

2. Methods

The mPower dataset holds records of participants vocalising the phoneme "aaah" [3]. We arranged the classes similar to Tracy et al. [2], but not considering UPDRS (Unified Parkinson's Disease Rating Scale) scores. We designed the stratified sampling to align datasets to a desired distribution of attributes by iteratively removing samples [4]. Our goal was to balance gender and classes, and match the age-distribution (10-years bins). Trade-off steps kept a minimal fraction of the dataset. To control for accuracy loss due to reduced train set size, we additionally assembled sets from random subsampling.

[1] Corresponding Author: Alexander Brenner, E-mail: alexander.brenner@uni-muenster.de.

We used the following feature sets from Surfboard [5] and OpenSmile [6]: 1) Surfboard PD, 2) eGeMAPS, 3) AVEC, and 4) ComParE. We tested two classifiers comparable to methods from previous studies, the Support-Vector-Machine (SVM) and CatBoost.

3. Results

We report accuracy for each combination of data and feature set using 5-fold cross-validation (Table 1). Results for CatBoost and further metrics are in the supplements [4].

Table 1. Mean classification accuracy (in %) on mPower subsets using the SVM and 5-fold CV (± std).

Setting	Surfboard PD	ComParE	AVEC	eGeMAPS
Complete set, CV	83.17 (0.65)	82.41 (0.18)	83.14 (0.47)	76.50 (0.67)
Complete set, grouped CV	71.29 (1.50)	71.45 (1.75)	70.96 (2.02)	65.43 (3.02)
Stratified sampling (50%)	61.35 (2.82)	63.14 (2.30)	62.04 (1.94)	57.99 (2.23)
Stratified sampling (20%)	56.20 (5.29)	57.12 (5.57)	56.53 (6.38)	52.61 (5.59)
Random sampling (50%)	70.19 (2.04)	70.35 (2.11)	69.99 (2.56)	65.14 (1.86)
Random sampling (20%)	67.79 (4.48)	67.41 (2.47)	67.10 (3.09)	63.64 (2.96)

4. Discussion

We observed high accuracy comparable to previous works. Grouping by individuals clearly reduced the scores, confirming identity confounding. When balancing age and gender distribution, we observed a further drop. While smaller sample size slightly decreased accuracy, random sampling outperformed stratified sampling. Although our results are limited to certain methods and a simple phonation task, we showed that it is crucial to consider comparable control groups when assessing performance.

5. Conclusion

We investigated smartphone-recorded phonation in PD detection using ML and the potential evaluation bias on unbalanced data. Many features show promising results on the mPower dataset, but are limited when balancing classes, gender and age. To prevent pitfalls in future analyses, we propose the use of stratified sampling and grouped CV.

References

[1] Mei J, Desrosiers C, Frasnelli J. Machine learning for the diagnosis of Parkinson's disease: a review of literature. Frontiers in aging neuroscience. 2021 May 6;13:633752.

[2] Tracy JM, Özkanca Y, Atkins DC, Ghomi RH. Investigating voice as a biomarker: deep phenotyping methods for early detection of Parkinson's disease. Journal of Biomedical Informatics. 2020 Apr 1;104:103362.

[3] Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, Doerr M, Pratap A, Wilbanks J, Dorsey E, Friend SH. The mPower study, Parkinson disease mobile data collected using ResearchKit. Scientific data. 2016 Mar 3;3(1):1-9.

[4] Brenner A. Stratified Sampling, (2023). https://github.com/alex-bre/iterative_stratified_sampling

[5] Lenain R, Weston J, Shivkumar A, Fristed E. Surfboard: Audio feature extraction for modern machine learning, *ArXiv Prepr. ArXiv200508848*. (2020).

[6] Eyben F, et al. Opensmile: the munich versatile and fast open-source audio feature extractor. InProceedings of the 18th ACM international conference on Multimedia 2010 Oct 25 (pp. 1459-1462).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230080

Does Using a Stacking Ensemble Method to Combine Multiple Base Learners Within a Database Improve Model Transportability?

Cynthia YANG[a,1], Egill A. FRIDGEIRSSON[a], Jan A. KORS[a], Jenna M. REPS[a,b], Peter R. RIJNBEEK[a], Jenna WONG[c] and Ross D. WILLIAMS[a]
[a] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[b] *Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ, USA*
[c] *Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, USA*

Abstract. We investigated a stacking ensemble method that combines multiple base learners within a database. The results on external validation across four large databases suggest a stacking ensemble could improve model transportability.

Keywords. Clinical prediction model, external validation, stacking ensemble

1. Introduction

When developing a clinical prediction model, it is impossible to know beforehand which modeling method is suitable for a particular prediction task. Additionally, prediction performance often drops when a model is transported to another database [2]. A stacking ensemble provides the opportunity to combine predictions from multiple base learners [1]. In [2], the authors show that ensembles combining lasso logistic regression models trained on different databases transported better than single database models. Our aim is to investigate whether using a stacking ensemble method to combine different base learners within a single observational health database improves model transportability.

2. Methods

We developed models using the Observational Health Data Sciences and Informatics (OHDSI) Patient-Level Prediction framework [3]. We used three large claims databases from the United States of America and one large electronic health record database from

[1] Corresponding Author: Cynthia Yang, c.yang@erasmusmc.nl. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Germany with data mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM): IBM MarketScan® Commercial Database (CCAE), IBM MarketScan® Multi-State Medicaid Database (MDCD), IBM MarketScan® Medicare Supplemental Database (MDCR), and IQVIA Disease Analyser Germany EMR (IQVIA Germany). Each site obtained institutional review board approval for the study or used de-identified data. We investigated 21 prediction tasks predicting 21 different outcomes of interest [3]: "Amongst a target population of patients with pharmaceutically treated depression, which patients will develop <the outcome> during the 1-year time interval following the start of the depression treatment?". We sampled an initial study population of 100,000 patients from each database. For each prediction task and database, we developed a stacking ensemble consisting of 3 different base learners (lasso logistic regression, random forest, and XGBoost) and a single meta-learner (logistic regression). A random subset of 75% of the patients was used for training and hyperparameter tuning of the base learners and the predictions of the base learners on the remaining 25% of the patients were used to train the meta-learner. To assess model transportability, we externally validated each model across the other three databases and evaluated the area under the receiver operating characteristic curve (AUC).

3. Results

On average, the stacking ensemble resulted in small positive AUC differences (Figure 1). All differences were significantly different from zero ($p < 0.05$), with the exceptions of lasso logistic regression in CCAE and IQVIA Germany, and XGBoost in CCAE.

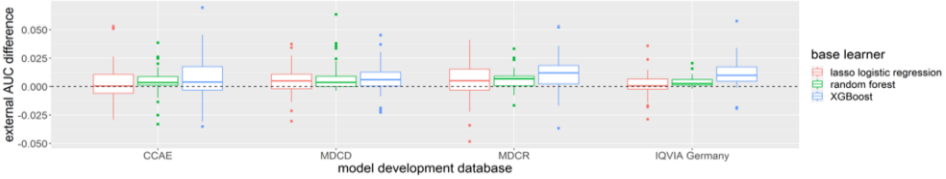


Figure 1. $External\ AUC\ difference = ensemble\ AUC - base\ learner\ AUC$. E.g., for CCAE, the blue box plot shows the ensemble AUC minus the XGBoost AUC across all outcomes and across all other databases.

4. Discussion and conclusion

The results suggest that using a stacking ensemble method can improve transportability of prediction models. However, the AUC differences were generally small, with the largest gain in AUC found for using the stacking ensemble instead of XGBoost alone. Future research may consider a larger set of base learners and different meta-learners.

References

[1] Wolpert DH. Stacked generalization. Neural Netw. 1992 Jan;5(2):241–59.
[2] Reps JM, et al. Learning patient-level prediction models across multiple healthcare databases: evaluation of ensembles for increasing model transportability. BMC Med Inform Decis Mak. 2022 May;22(1):1-4.
[3] Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018 Aug;25(8):969-75, doi: 10.1093/jamia/ocy032.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230081

Implementation of the Regulation on the European Health Data Space (EHDS) in Croatia

Emanuel BRAĐAŠEVIĆ[a,1]
[a] *Croatian Institute of Public Health, Zagreb, Croatia*
ORCiD ID: Emanuel Brađašević https://orcid.org/0000-0001-9398-7403

Abstract. Since the European Commission has published a Proposal for a Regulation – The European Health Data Space, Croatia is actively working on the implementation. Public sector bodies including the Croatian Institute of Public Health, Ministry of Health and Croatian Health Insurance Fund play a key role in this process. The establishment of a Health Data Access Body is the main challenge in this effort. Potential challenges and obstacles in this process and projects that follow the efforts are described in this paper.

Keywords. European Health Data Space, Health Data Access Body, health data

1. Introduction

On May 2022 the European Commission has published a Proposal for a regulation – The European Health Data Space. [1] This regulation is a follow-up and first domain-specific regulation on The European strategy for Data and will be an integral part in the creation of a European Health Union. [2] As a member state of the European Union, Croatia has the obligation to implement the Regulation on national level. The aim of this paper is to identify the key stakeholders responsible for the implementation in Croatia, map the projects that are a part of the implementation process and to list potential challenges and obstacles in the implementation.

2. Stakeholders

Public sector bodies play a key role in the implementation of the EHDS Regulation, mainly, Croatian Institute of Public Health (CIPH), Ministry of Health (MIZ) and Croatian Health Insurance Fund (CHIF). The Ministry of Health plays a role by directly implementing legislation on a national level, while CHIF is a data holder. CIPH is also a data holder and processor and is already responsible for the maintenance of 16 national registries related to health data. In the framework of this Regulation, CIPH is devoted to create the first Health Data Access Body (HDAB) as a part of its institution through projects with MIZ and CHIF as partners.

[1] Corresponding author: Emanuel Brađašević, Croatian Institute of Public Health, Rockfellerova 12, 10000 Zagreb, Croatia; E-mail: emanuel.bradasevic@hzjz.hr

3. Projects

The European Commission encourages member states to implement the Regulation with financial and logistic support through Direct Grants and Joint Actions. Croatia has successfully performed the project TEHDAS - Joint Action Towards the European Health Data Space as part of this effort. Currently CIPH is taking part in HealthData@EU - European health data space pilot for secondary use of health data with the aim to test and evaluate the IT infrastructure on five concrete cases from which Croatia takes part in three. Data sources used in these pilots include the National Public Health Information System (NAJS) linked to databases held by CHIF and Tax administration of the Ministry of Finance. Croatia will also apply for Direct grants to Member states for setting up services by health data access bodies - secondary use of health data (HDAB), Direct grants to Member states for expansion of MyHealth@EU Digital Service Infrastructure (eHDSI) with new services and to more member states and Direct grants to Member States' authorities: preparatory actions for a European Health Data Space; primary use of data (for healthcare) and reuse of data.

4. Challenges and Obstacles

The biggest challenge for Croatia is the establishment of the first and only planned Health Data Access Body (HDAB) led by the CIPH. Human and financial resource limitations in the establishment of such a new part of the institution are the main concern. Technical infrastructure building, achievement of interoperability with other member states and connection of data stored internally in hospitals adds to the complexity of the project. Croatia has not yet built a metadata catalogue which is also obligatory in the achievement of interoperability with other states. Since MIZ and CHIF are working in partnership with CIPH on most of the activities, they face the same problems. Communication and harmonization of activities between those three public sector bodies in Croatia is essential in achieving the goal of successful implementation of the EHDS Regulation.

5. Conclusion

Three public sector bodies in Croatia are actively working on the implementation of the Regulation on the EHDS: Ministry of Health, Croatian Institute of Public Health and Croatian Health Insurance Fund. The establishment of the first and only Health Data Access Body as part of CIPH is the biggest task to be fulfilled. Financial and human resource limitations are among the most concerning obstacles that all institutions face.

References

[1] European Health Data Space [Internet]. Public Health. [cited 2023 Jan 10]. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

[2] REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space [Internet]. EUR-lex. EUROPEAN COMMISSION; 2022 [cited 2023 Jan 10]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230082

Mapping the SPHN Dataset to FHIR

Sophie Anne Inès KLOPFENSTEIN[a,b,1], Sylvia THUN[a], Katrin CRAMERI[c] and Caroline STELLMACH[a]
[a] *Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Germany*
[b] *Charité – Universitätsmedizin Berlin, Germany*
[c] *Swiss Institute of Bioinformatics, Switzerland*
ORCiD ID: S. A. I. Klopfenstein https://orcid.org/0000-0002-8470-2258,
S. Thun https://orcid.org/0000-0002-3346-6806,
K. Crameri https://orcid.org/0000-0003-3656-3457,
C. Stellmach https://orcid.org/0000-0001-6798-8533

Abstract. Several European health data research initiatives aim to make health data FAIR for research and healthcare, and supply their national communities with coordinated data models, infrastructures, and tools. We present a first map of the Swiss Personalized Healthcare Network dataset to Fast Healthcare Interoperability Resources (FHIR®). All concepts could be mapped using 22 FHIR resources and three datatypes. Deeper analyses will follow before creating a FHIR specification, to potentially enable data conversion and exchange between research networks.

Keywords. Standards, Interoperability, FHIR, SPHN, RDF, Datasets

1. Introduction

The Swiss Government funds the Swiss Personalized Health Network (SPHN) to foster the ability to share health data for research and ultimately enable personalized health, as part of their eHealth strategy. The SPHN and German Medical Informatics Initiative (MII) share these goals and create, validate and implement necessary data infrastructures and tools [1, 2]. The "SPHN Semantic Interoperability Framework" includes a dataset composed of fully defined, combinable informational units (concepts), that can be bound to semantic standards and value sets. A Resource Description Framework (RDF) schema is used as exchange format [3]. The MII chose the syntactic standard Fast Healthcare Interoperability Resources (FHIR), building on resources leveraging semantic standards to define content and structure of healthcare concepts [4]. Envisaging to enable data exchange between research networks, we present a first analysis of a SPHN to FHIR map.

2. Methods

We filtered the SPHN dataset (version 2022.1) [5] on the columns "active status (yes/no)" and "concept or concept compositions or inherited". Two authors mapped independently

[1] Corresponding Author: Sophie A. I. Klopfenstein, Berlin Institute of Health and Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany; E-mail: sophie.klopfenstein@charite.de.

the active concepts to FHIR (v4.0.1) based on their descriptions and noted information from the specification [4]. All authors discussed both maps to derive a consolidated map.

3. Results and Discussion

All active SPHN concepts could be mapped to FHIR using 22 unique FHIR resources (59/63 SPHN concepts), and three datatypes (4/63 SPHN concepts): Most were mapped to the "Observation" (20/63), "Patient" (6/63), "Condition" and "Procedure" (both 5/63) resources. Five concepts could not be unambiguously mapped to a single FHIR resource. A more complex modeling is necessary to fully represent 22 SPHN concepts, requiring several FHIR elements used in combination or calculations. For some SPHN concepts, the exact datatype cannot be defined in the map as it might depend on the context [6].
All SPHN concepts could be represented using FHIR resources and datatypes, but several SPHN concepts need a complex representation due to their genericness. By using RDF, the dataset does not follow a data model-based approach, allowing the framework to be applied in a broad way, but impeding direct mapping to FHIR. SPHN aims to cooperate with diverse research communities by developing maps and conversion of the RDF schema [3]. To complete the map and ease the creation of a FHIR specification for the dataset, all other SPHN elements beside "concepts" must be mapped to FHIR, cardinalities evaluated, value sets compared and adapted, and required FHIR elements that currently do not have an equivalent in the SPHN dataset be added. Preexisiting FHIR profiles should also be considered during specification. To foster data exchange and improve health research cross-border, an in-depth comparison of the content covered in MII's core data set modules and the SPHN dataset should be performed as well.

4. Conclusion

Mapping SPHN to FHIR is overall possible. Deeper analyses based on example data should follow before profiling the SPHN dataset in FHIR. Framing the SPHN dataset with a FHIR specification would aid the seamless exchange of research data generated within the SPHN and MII. A conversion tool could be helpful on the long term to enable data exchange between the German and Swiss health research communities.

References

[1] Österle S, Touré V, Crameri K. The SPHN Ecosystem towards FAIR data. CEUR Workshop Proceedings. 2021. p. 19-28
[2] Medical Informatics Initiative. About the initiative [Internet]. Available from: https://www.medizininformatik-initiative.de/en/about-initiative
[3] Gaudet-Blavignac C, et al. A National, Semantic-Driven, Three-Pillar Strategy to Enable Health Data Secondary Usage Interoperability for Research Within the Swiss Personalized Health Network: Methodological Study. JMIR Med Inform. 2021 Jun 24;9(6):e27591.
[4] HL7 International. HL7 FHIR Release 4 - Welcome to FHIR® [Internet]. 2019. Available from: https://www.hl7.org/fhir/R4/
[5] Swiss Personalized Health Network (SPHN). SPHN Dataset Release [Internet]. 2022. Available from: https://sphn.ch/document/sphn-dataset/
[6] BIH-CEI/sphn-to-fhir-map [Internet]. Available from: https://github.com/BIH-CEI/sphn-to-fhir-map

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230083

Assessing Quality of Life Using FHIR – How to Combine Patient Reported Outcome with Patient Generated Data for Better Compliance

Chantal N.L. BEUTTER[a,1], Patrick WERNER [a], Stefan SIGLE[a], Uwe Marc MARTENS[a,b] and Christian FEGELER[a,c]
[a] *MOLIT Institute gGmbH, Heilbronn, Germany*
[b] *SLK Kliniken GmbH, Heilbronn, Germany*
[c] *University of applied Science Heilbronn, Heilbronn, Germany*

Abstract. Quality of life (QoL) is affected by environmental influences and varies between patients. A combined measurement through Patient Reported Outcomes (PROs) and Patient Generated Data (PGD) may enhance the detection of QoL impairments by a longitudinal survey. Leveraging different approaches of QoL measurement techniques, the challenge is to combine data in a standardized, interoperable way. We developed an app (*Lion-App*) to semantically annotate data from sensor systems as well as PROs to be merged in an overall analysis of QoL. A FHIR implementation guide was defined for a standardized assessment. To access sensor data the interfaces of Apple Health or Google Fit are used instead of integrating various provider directly into the system. Since QoL cannot be collected exclusively via sensor values, a combination of PROs and PGD is necessary. PGD enable a progression of QoL which offers more insight into personal limitations whereas PROs give insight about personal burden. The use of FHIR enables structured exchange of data while personalized analyses might improve therapy and outcome.

Keywords. Quality Of Life, FHIR, Interoperability, Patient Reported Outcome, Patient Generated Data

1. Introduction

To detect personal deviations in Quality of Life (QoL) a longitudinal survey and good compliance is necessary. This may be achieved by combining subjective Patient Reported Outcomes (PROs) with objective Patient Generated Data (PGD). Merging two measurement techniques, standardization and therefore interoperability are important [1].

[1] Corresponding Author: Beutter, C.N.L., MOLIT Institute gGmbH, Im Zukunftspark 10, 74076 Heilbronn, Germany, E-Mail: chantal.beutter@molit.eu.

2. Methods

Currently measurement techniques of QoL have the limitation that the absence of behavioral reference and no longitudinal monitoring may exhibit a response bias [2].

Therefore, an app (*Lion-App*) was conceptualized and implemented for a continuous measurement of QoL. PROs and PGD needed to be uniformly structured to be combined.

3. Results and Discussion

In Lion-App QoL is gathered through a sensor-based survey of daily activities and PROs about symptoms. All is structured through HL7 FHIR as it utilizes LOINC or SNOMED codes to further use data semantically interoperable in computations or health care delivery. Therefore, data from PROs and PGD can be matched by the same semantic annotation for a joint analysis independent of their data source. A FHIR implementation guide was developed to define resources for a standardized assessment.

Our concept is based on the use of private devices for a universal applicability across different hardware with supporting the operation systems of Android or iOS. Instead of implementing interfaces to various providers, Apple Health and Google Fit have been prioritized to access and ingest sensor data. User centered development of the app showed positive results for applicability and acceptance of the app [3]. A basic requirement for measuring QoL through PROs and PGD is a compliant patient. A purely objective survey through sensors (PGD) can extend the survey by behavioral reference, but it cannot replace the subjective assessment of the personal burden (PROs). For this reason, a combination is necessary to enable reliable, longitudinal QoL tracking. With the automated behavioral analysis in combination with PROs and mapping of the data to QoL, trends can be identified at an early stage. To use data in medicine, it must be semantically coded in a uniform way, for example by using HL7 FHIR. To date, no such systems are known to collect QoL in oncology, which is why it was decided to publish the structure of our approach in an implementation guide as the concept may be reused by other members of the scientific community.

4. Conclusion

In summary, it is technically possible to implement a system for the combined measurement of QoL. By using the communication standard of HL7 FHIR, mobile collected PROs and PGD can be combined and further used in health care systems.

References

[1] Bakouny Z, Patt DA. Machine learning and real-world data: More than Just Buzzwords. JCO Clinical Cancer Informatics. 2021 Aug;5:811-3.

[2] S. M. Goldberg u. a., „The roles of study setting, response bias, and personality in subjective memory complaints of cognitively normal older adults“, Int. Psychogeriatr., Bd. 33, Nr. 7, S. 665–676, Juli 2021,

[3] C. N. L. Beutter, K. Zeller, U. M. Martens, B. Pfleiderer, und C. Fegeler, „Development of a mobile application to assess quality of life of cancer patients: Results of an iterative investigation and usability testing using a user-centered approach (Preprint)“, JMIR Cancer, preprint, Dez. 2022. doi: 10.2196/preprints.44985.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230084

Portal of Medical Data Models: Application in Federated Data Capture

Matthias GANZINGER[a,1], Max BLUMENSTOCK[a], Christian NIKLAS[a], and Martin DUGAS[a]
[a]*Institute for Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany*

Abstract. So far, the portal for medical data models allows its users to download medical forms in a standardized format. Importing data models into electronic data capture software involved a manual step of downloading and importing the files. Now, the portal was enhanced with a web services interface to allow electronic data capture systems to automatically download the forms. This mechanism can be used in federated studies to ensure that all partners are working with identical definitions of study forms.

Keywords. Medical Data Models, Case Report Forms, Federated Data Capture

1. Introduction

The portal for medical data models (MDM) is a repository that provides access to standardized medical data models and related resources [1]. Currently, it contains more than 24,000 forms from different medical areas including health record forms or case report forms (CRF) from clinical trials. Typically, MDM users download forms manually using a web browser in order to import the form definitions into an electronic CRF (eCRF) system of their choice later on.

In multicentric studies the participating sites often agree on a single installation of CRF software that is made available over the Internet to all partners. However, since potentially sensitive patient data are transferred to institutions without a treatment context in this scenario, a central system has drawbacks in terms of data protection. A way to mitigate this problem is to install an instance of the eCRF software locally at each site. In this federated scenario it becomes essential to keep forms definitions synchronized across all sites to ensure generation of consistent high-quality data sets.

Federation of data management systems is a current topic in Medical Informatics research. For example, Liu et al. describe a federated architecture on medical data security [3]. Moshawrab et al. report on federated learning approaches for disease prediction [4].

We propose a novel interface for MDM that allows automated machine-to-machine access of form definitions within the portal. No personal data are exchanged in this step. In addition to individual systems, this interface can be used for metadata synchronization in federated data capture environments.

[1] Corresponding Author: Matthias Ganzinger, IMI UKHD, Im Neuenheimer Feld 130.3, 69210 Heidelberg, Germany; E-mail: matthias.ganzinger@med.uni-heidelberg.de

2. Methods

The primary data format of MDM is Clinical Data Interchange Standards Consortium Operational Data Model (CDISC ODM), but export in other formats, such as Health Level Seven Fast Healthcare Interoperability Resource (HL7 FHIR) questionnaire is possible as well [2]. MDM portal is accessible via Hypertext Transfer Protocol Secure (HTTPS) for users. Consequently, a Representational State Transfer (REST) interface for machine access to ODM metadata was added. This interface allows to reference and download a specific form definition using the MDM identifier from the portal.

3. Results

For a proof of concept, MDM is now equipped with a REST interface. As a first consumer application of the interface we enhanced the electronic data capture (EDC) software openEDC [5] with a REST client interface. By entering the MDM identification number of a data model in openEDC, the corresponding form is automatically download via REST and made available as CRF. The software was tested in a federated test setting: four instances of openEDC located in German university hospitals referenced the same forms definition on MDM. Then, data for fictitious patients were manually entered into the openEDC instances to simulate a multicentric study without the restrictions introduced by data protection end ethics requirements for real patient data. Finally, study data were exported and sent to the study center and finally combined for analysis.

4. Discussion

We have demonstrated that MDM can be used as a central repository for EDC in federated scenarios. Using a single source for metadata definitions at all participating sites ensures that forms are in sync for the complete study. While the REST interface was tested with openEDC, its specification and access will be made available to the general public. Manufacturers of other EDC programs are encouraged to adopt the interface in their systems. To accommodate potential needs for adaption in specific study settings, we will investigate how customizing and synchronizing forms for specific study needs beyond the standard forms can be supported by MDM as well.

References

[1] Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P et al. Portal of medical data models: information infrastructure for medical research and healthcare. Database (Oxford). 2016;2016, doi: 10.1093/database/bav121.

[2] Doods J, Neuhaus P, Dugas M. Converting ODM Metadata to FHIR Questionnaire Resources. Stud Health Technol Inform. 2016;228:456–60, doi: 10.3233/978-1-61499-678-1-456.

[3] Liu X, Zhao J, Li J, Cao B, Lv Z. Federated Neural Architecture Search for Medical Data Security. IEEE Trans Ind Inf. 2022;18(8):5628–36, doi: 10.1109/TII.2022.3144016.

[4] Moshawrab M, Adda M, Bouzouane A, Ibrahim H, Raad A. Reviewing Federated Machine Learning and Its Use in Diseases Prediction. Sensors (Basel, Switzerland). 2023;23(4), doi: 10.3390/s23042112.

[5] Greulich L, Hegselmann S, Dugas M. An Open-Source, Standard-Compliant, and Mobile Electronic Data Capture System for Medical Research (OpenEDC): Design and Evaluation Study. JMIR Med Inform. 2021;9(11), doi: 10.2196/29176.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230085

The DELPHI Library: Improving Model Validation, Transparency and Dissemination Through a Centralised Library of Prediction Models

Ross D. WILLIAMS[a,1], Sicco DEN OTTER[a], Jenna M. REPS[a,b] and Peter R. RIJNBEEK[a]
[a] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[b] *Janssen Research and Development, Raritan, NJ, USA*

Abstract. The Deposit, Evaluate and Lookup Predictive Healthcare Information (DELPHI) library provides a centralised location for the depositing, exploring and analysing of patient-level prediction models that are compatible with data mapped to the observational medical outcomes partnership common data model.

Keywords. Prediction Model, Interoperability, communication

1. Introduction

Over the past decade there has been a rapid increase in the number of published clinical prediction models [1], but no similar rise in use within clinical settings. This usage gap is due to multiple reasons including insufficient reporting [2], non-publication of full models, inadequate testing [3] and lack of trust from clinical stakeholders. To improve reporting and model sharing we created the Deposit, Evaluate and Lookup Predictive Healthcare Information (DELPHI) library. A centralised repository where users can upload Patient-Level prediction (PLP) models, performance statistics and complete information on all model design choices. This enables others to explore results, replicate model development and externally validate models downloaded from DELPHI.

2. Methods

We used the interoperability provided by a network of databases mapped to a common data model and the standardised analytical pipeline provided by the PatientLevelPrediction R package [4] to create a database for models. A graphical user interface provides interested parties with a simple method of searching for and evaluating

[1] Corresponding Author: Ross D. Williams, E-mail: r.williams@erasmusmc.nl

prediction models and their validations, as well as uploading and downloading models and results. This is provided as a web application.

3. Results and Discussion

The library is available at: https://delphi.ohdsi.org/

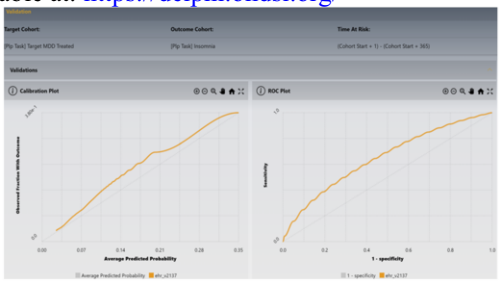


Figure 1. The validation tab detailing the ROC plot and calibration plot along with performance metrics to analyse the selected prediction model.

Researchers are encouraged to explore and upload models to contribute to the library. Figure 1 shows the validation exploration page. The models are searchable by patient cohorts, algorithm type, database used, and developing researcher. Once a model is found, all key model and performance information will be viewable. DELPHI provides a dynamic results exploration environment that is unavailable in other repositories [5]. The main purpose of PLP models is to influence clinical practice. In order to do this the level of trust in the modelling process and the models themselves needs to be improved. Currently, prediction models are spread throughout the scientific literature and often poorly reported. The DELPHI library provides a centralised location to store models and results and makes accessible everything that is needed to assess and implement PLP models. The interactivity of DELPHI provides an extra dimension than is in other similar efforts [5].

4. Conclusions

A centralised, standardised model repository should help improve the searchability, external validation and assessment of prediction models. It is hoped this leads increased trust in and usage of models in clinical practice.

References

[1] Yang C, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. J Am Med Inform Assoc. 2022;29(5):983-9.
[2] Moons KG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-73.
[3] Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245-7.
[4] Reps JM, et al. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018.
[5] Wessler BS, Paulus J, Lundquist CM, Ajlan M, Natto Z, Janes WA, et al. Tufts PACE Clinical Predictive Model Registry: update 1990 through 2015. Diagn Progn Res. 2017;1:20.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230086

Continuity of Patient Information to Palliative Care

Anne KUUSISTO[a,b,1], Kaija SARANTO[c], Päivi KORHONEN[a,d] and Elina HAAVISTO[a,e,f]
[a]*University of Turku*
[b]*The wellbeing services county of Satakunta*
[c]*University of Eastern Finland*
[d]*Turku University Hospital*
[e]*University of Tampere*
[f]*Tampere University Hospital*

ORCiD ID: Anne Kuusisto https://orcid.org/0000-0002-9950-1985, Kaija Saranto https://orcid.org/0000-0002-3195-1955, Päivi Korhonen https://orcid.org/0000-0002-0244-4890, Elina Haavisto https://orcid.org//0000-0002-9747-1428

Abstract. The purpose of this study was to describe the sharing of knowledge and information to palliative care in terms of information content, information structure and information quality by means of Advance Care Planning (ACP). This study used a descriptive qualitative study design. Purposively selected nurses, physicians and social workers working in palliative care in Finland took part in thematic interviews in five hospitals in three hospital districts in 2019. The data (n = 33) were analyzed by means of content analysis. The results demonstrate the evidence-based practices of ACP in terms of information content, structure and quality. The results of this study can be utilized in the development of sharing knowledge and information and as the basis in the development of an ACP instrument.

Keywords. Information dissemination, knowledge, continuity of patient care, health care professionals, hospital information systems, electronic health records

1. Introduction

The continuity of patient information refers to how well patients' health information accompanies them between different treatment settings and service providers [1,2]. Knowledge sharing is particularly important in palliative care, which refers to an approach that seeks to alleviate suffering in patients with a life-threatening illness [3]. Advance Care Planning (ACP) would constitute part of an individual care and support care planning. This kind of advance care plans can be recorded in a patient's electronic health record [4]. A scoping review shows deficiencies in the continuity of care in terms of transfer of information concerning patients who are in need of palliative care [5].

The purpose of this study is to describe the sharing of knowledge and information to palliative care in terms of information content, information structure and information quality by means of Advance Care Planning.

[1] Corresponding Author: Anne Kuusisto, E-mail: anmakuu@utu.fi.

2. Methods

This study used a descriptive qualitative study design. The study data were collected from purposively selected nurses (n = 18), practical nurses (n = 5), physicians (n = 5) and social workers (n = 5) who took part in thematic interviews in five hospitals in three hospital districts in Finland in 2019. Interviews with nurses were mainly conducted in focus groups, whereas physicians and social workers were mainly interviewed individually, one at a time, because there would not have been enough of them to form focus groups in individual organizations. The data (n = 33) were analyzed using content analysis. University Research Ethics Committee granted ethical approval (15/2019).

3. Results

All the interviewees were experts with long work experience in palliative care. They had an average of 17 years of work experience in health care and an average of six years' of work experience in palliative care. The results show the evidence-based practices of the ACP in terms of information content, information structure and information quality from the multi-professional perspective. According to this study, the quality of information is multi-dimensional and by using its different perspectives in other words availability, informativeness and usability, it is possible to evaluate the quality of patient information coming to palliative care.

4. Discussion

Further development of documenting ACP is needed, particularly when it comes to communication practices, as well as Hospital Information Systems in terms of interoperability.

5. Conclusion

The results of this study can be utilized in the development of sharing knowledge and information and as the basis in the development of an ACP instrument.

References

[1] Haggerty Jl, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R. Continuity of care: a multidisciplinary review. BMJ 2003 Nov;327(7425):1219-21. doi: 10.1136/bmj.327.7425.1219.
[2] Ohno-Machado L, Séroussi B, editors. Contents of informational and management continuity of care. Proceedings of the 17th World Congress of Medical and Health Informatics (Medinfo); Aug 25-Aug 30; Lyon France. Anne Kuusisto, Paula Asikainen, Kaija Saranto; 264:669-673. doi: 10.3233/SHTI190307.
[3] Connor SR, editor. Global Atlas of Palliative Care. 2nd ed. London, UK: Worldwide Hospice Palliative Care Alliance; 2020.
[4] NHS England. Advance care planning. What is advance care planning? https://www.england.nhs.uk/.
[5] Kuusisto A, Santavirta J, Saranto K, Korhonen P, Haavisto E. Advance care planning for patients with cancer in palliative care: A scoping review from a professional perspective. J Clin Nurs 2020 Jul;29(13-14):2069-2082. doi: 10.1111/jocn.15216.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230087

Secure Sharing of Health-Related Data: Research Description of the VINTER, DELFIN, and HEIDA Projects

Helena M. LINGE[a,1] and Rickard BRÄNNVALL[a]
[a]*RISE Research Institutes of Sweden, Sweden*

Abstract. The need for secure and integrity-preserved data sharing has become increasingly important in the emerging era of changed demands on healthcare and increased awareness of the potential of data. In this research plan, we describe our path to explore the optimal use of integrity preservation in health-related data contexts. Data sharing in these settings is poised to increase health, improve healthcare delivery, improve the offering of services and products from commercial entities, and strengthen healthcare governance, all with a maintained societal trust. The HIE challenges relate to legal boundaries and to the importance of maintaining accuracy and utility in the secure sharing of health-related data.

Keywords. Artificial Intelligence, GDPR, Sensitive Data, Privacy Preservation

1. Introduction and Presentation of the Projects

Technological advancements, resource-related insights, and changing demands on healthcare have made it evident that health-related data is a valuable asset. Data-driven patient outcome prediction, treatment planning, and resource optimization have great potential for improving health care. Data sharing may support awareness and compliance with treatments, suggest novel treatments, and speed up the resolution of complications or disease. With life-long chronic diseases, e.g., type I diabetes, the intersections between health care providers, the individual, and commercial entities, already constitute vibrant areas where values derived from data are created and exchanged. Sharing healthcare data between providers and utilizing it outside of traditional healthcare settings holds great potential. However, compliance with regulations like GDPR in the EU and HIPAA in the US, as well as the Cloud Act, present interoperability challenges that must be addressed. A broader adaptation of privacy-by-design may be necessary to maintain social trust in the new digital analytical health services as they rely on the sharing of data.

VINTER: In Spring 2022, a team from RISE won the Vinter innovation challenge hosted by the Swedish innovation authority Vinnova. The solution was based on homomorphic encryption, an emerging technology for confidential computing that makes possible the processing of encrypted data without access to the secret key [1]. The winning entry described analyses of blood glucose values from individuals with type I diabetes by a digital service using confidential computing. It proposed a software infrastructure building block that works as a data intermediary between the individual,

[1] Corresponding Author: Helena M. Linge, RISE, E-mail: helena.linge@ri.se

the health care provider, and a third-party service [2]. *HEIDA:* The VINTER-winning team was consequently awarded funds to develop example software components using homomorphic encryption and related technologies and to investigate business models for their implementation. This project focuses on 1) privacy-by-design solutions that support third-party cloud-based digital services and 2) the support of public sector efforts in using federated learning. The quest is pursued together with two public sector entities in western Sweden responsible for health care. Homomorphic encryption, alongside federated learning, was listed as a promising technology in Vinnova's recent report on secure data sharing [3]. *DELFIN:* In late spring 2022, Vinnova launched a call for a preparatory phase of building a large-scale demonstrator for system change in using health-related data. The funding agency selected seven proposals, including the DELFIN project led by RISE and supported by a large healthcare provider, a commercial entity, and two patient organizations. The demonstrator aims to exemplify the value of a unified infrastructure where data can be accessed and analyzed by many different parties, such as commercial entities, healthcare providers, individuals, and academic researchers. The initial focus relates to foot ulcers, a preventable complication of diabetes identified to benefit [4] from more efficient sharing of medical examination data including images. The demonstrator aims to be generalizable to other medical settings.

2. Discussion and Conclusions

In the long-term we believe that advanced privacy preservation technologies increase trust in digital society and thereby enable launches of applications previously not feasible due to legal considerations. This benefits both the individual citizen and the innovator by providing alternatives and opportunities. The current work aims at exploring not only the technical circumstances but also how, for example, behaviors, policies, regulations, infrastructure, and markets need to be changed for data sharing with integrity preservation to be practiced, perhaps in an automated fashion in the future. Here, we find inspiration in the Trusted Research Environments model [5] from Great Britain, which describes Five Safes: Safe people, Safe projects, Safe settings, Safe outputs, Safe data. The activities aim for a broader understanding of positioning and use of integrity-preserving technologies in order for society to take full advantage of health data. With the Vinnova Vinter challenge as a starting point, we pursue several paths to determine optimal use of integrity preservation technologies in contexts where data sharing may create value while maintaining personal integrity and societal trust.

References

[1] Gentry C. Computing arbitrary functions of encrypted data. Communications of the ACM. 2010 mar.
[2] Brännvall R, et al. Homomorphic encryption enables private data sharing for digital health: winning entry to the Vinnova innovation competition Vinter 2021-22. In: Pr. of Swedish AI Symposium; 2022.
[3] Vinnova, Sweden's innovation agency. Slutrapporten i regeringsuppdraget Uppdrag att kartlägga behov av utvecklingsinsatser för datadelning. DNR 2021-04320; 2022.
[4] Tang UH, et al. The D-Foot, for prosthetists and orthotists, a new eHealth tool useful in risk classification and foot assessment in diabetes. Foot and Ankle Online Journal. 2017;10(2).
[5] UK Health Data Research Alliance and NHSX, Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems, Zenodo, 2021.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230088

Improving Healthcare Quality with an LHS

Ricardo L. SANTOS[a,1] and Ricardo CRUZ-CORREIA[b]
[a] *PhD Student Health Data Science Faculty of Medicine University of Porto, Portugal*
[b] *Auxiliary Professor Faculty of Medicine University of Porto, Portugal*
ORCiD ID: Ricardo L. Santos https://orcid.org/0000-0001-6017-8255
Ricardo Cruz-Correia https://orcid.org/0000-0002-3764-5158

Abstract. The Learning Health System (LHS) is an important tool to help healthcare professionals solve problems by collecting, analyzing, interpreting and comparing health data, with the objective of helping patients make the best decision based on their own data, given the best evidence available. [1]. We believe partial oxygen saturation of arterial blood (SpO2) and related measurements and calculations can also be candidates for predictions and analysis of health conditions. We intend to build a Personal Health Record (PHR) that can exchange data with Electronic Health Records (EHRs) from hospitals, propose enhanced self-care, seek a support network, or look for healthcare assistance, (primary care or emergency service).

Keywords. Learning Health System, Interoperability, Vital signs, Personal Health Record

1. Introduction and Problem Description

This poster is a preview of a thesis that will be submitted to the Health Data Science PhD Program from the Faculty of Medicine of the University of Porto, Portugal. The thesis will be the first step in the long journey of developing a **streaming data platform to back the building of a Learning Health System (LHS)** in order to improve healthcare quality and its perceived value. The main idea behind LHS is the continuous improvement of outcomes through the generation and employment of knowledge during healthcare delivery by the use of informatics, science and enhancement of education, training and performance [1]. The LHS is an important tool to help healthcare professionals solve problems by collecting, analyzing, interpreting and comparing health data, with the objective of helping patients make the best decision based on their own data, given the best evidence available. **Clinical Decision Support Systems (CDSS)** are just one of the main purposes of an LHS. Interoperability: has the power to integrate the structure and semantics of health data and prevent loss of meaning of conceptual domains, knowledge in context and formal representation of data [2]. Besides data transportation (HL7 V2, V3 and FHIR), security (HTTPS) and structure and format (CDA, OMOP), health data standards involve vocabulary (SNOMED-CT), terminologies (LOINC) and classifications (ICD-10 and 11) used to **describe the real world in healthcare** [1]. **Vital signs**: as reported by Hirten, RP. [3] in 2020 regarding COVID-19 infections in healthcare professionals through **heart rates measured by smartwatches**, participants

[1] Corresponding Author: Ricardo L. Santos, E-mail: ricardolourencosantos@gmail.com.

wore Apple Watch and an app on their smartphones. We believe **partial oxygen saturation of arterial blood (SpO2)** and related measurements and calculations can also be candidates for predictions and analysis of health conditions, as so **body temperature** and others that are yet to become measurable through smart devices.

Literacy: self-care measures may be proposed to the patient through **self-learning**, and consequently avoid the clinical condition worsening and seeking help sooner, at a smaller economic and social cost, which can be measured by proper indicators, giving adequate data. This will help increase the number of patients that are able to self-manage and improve their health conditions through literacy and self-knowledge. Patients responsive to suggested education can reduce or avoid presential medical appointments for prescriptions (mainly reissuing), taking vital signs, simple orientations, feedback and follow-up with no or minor reportable events reducing the number of unnecessary medical appointments at primary care (and many times at emergency services), but also avoid treatments and hospitalizations, as the patients improve their health conditions.

2. Aim and Method

The cycle of care, research and knowledge discovery and its application into healthcare begins with the patient, as the history and the evolution of the pathophysiology of the disease itself (and its related signs and symptoms) are **translated into changes in vital signs (prior to lab and image exams)**, and those can be recorded into exchangeable, semantically interoperable and meaningful data.

3. Resutls

We intend to build a Personal Health Record (PHR) that can exchange data with Electronic Health Records (EHRs) from hospitals, propose enhanced self-care, seek a support network, or look for healthcare assistance, (primary care or emergency service).

4. Conclusions and Future Work

Still missing and to be addressed at the systematic review are GDPR and Medical Devices Regulation, as so is defining the unique identifier for each patient.

References

[1] Foley T, Horwitz L, Zahran R. Realising the potential of Learning Health Systems - The Learning healthcare Project. The Health Foundation, Newcastle University. May, 2021. Available at: https://learninghealthcareproject.org/publications/

[2] Chatterjee A, Pahari N, Prinz A. HL7 FHIR with SNOMED-CT to Achieve Semantic and Structural Interoperability in Personal Health Data: A Proof-of-Concept Study. Sensors. 2022 May 15;22(10):3756. doi: https://doi.org/10.3390/s22103756

[3] Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Longitudinal Physiological Data from a Wearable Device Identifies SARS-CoV-2 Infection and Symptoms and Predicts COVID-19 Diagnosis. Infectious Diseases (except HIV/AIDS); 2020 Nov [cited 2023 Jan 10]. doi: https://doi.org/10.1101/2020.11.06.20226803

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230089

The MeDaX Knowledge Graph Prototype

Judith A.H. WODKE[a,1], Lea MICHAELIS[b] and Ron HENKEL[a]
[a] *Department of Medical Informatics, University Medicine Greifswald, Germany*
[b] *Core Unit Data Integration Center, University Medicine Greifswald, Germany*
ORCiD ID: Lea Michaelis https://orcid.org/0000-0001-9691-2677

Abstract. Data sharing is sustainable for several reasons, including minimising economical and human costs or maximising knowledge gain. Still, reuse of biomedical (research) data is often hampered by the diverse technical, juridical, and scientific requirements for biomedical data handling and specifically sharing. We are building a toolbox for automated generation of knowledge graphs (KGs) from diverse sources, for data enrichment, and for data analysis. Into the MeDaX KG prototype, we integrated data from the core data set of the German Medical Informatics Initiative (MII) with ontological and provenance information. This prototype is currently used for internal concept and method testing only. In subsequent versions it will be expanded by including more meta-data and relevant data sources as well as further tools, including a user interface.

Keywords. Knowledge graphs, biomedical data, data enrichment, data reuse, open source, MeDaX

1. Introduction

The MII [1], aiming at digitisation of health care in Germany, follows a federated storage approach: Every German university clinic has set up a data integration center (DIZ) that, based on a common core data set (CDS) [2], provides digital solutions for biomedical data management. And while the CDS itself is standardised in HL7/FHIR [3] format, technical solutions provided for data storage, management, and analysis often are not.

KGs have proven suitable for representation of complex heterogeneous data [4,5]. Within the MeDaX project, we are building a toolbox for bio**Me**dical **Da**ta e**X**ploration. This includes innovative and efficient methods for data harmonisation and storage in KGs, for data enrichment, analysis, and retrieval. The presented first prototype serves internal testing purposes. With the entire project, apart from data FAIRification [7] and improving reuse opportunities for biomedical data, we aim at informing and empowering several stakeholders at the same time: medical personnel, scientists, and the public, including the patients the data might originate from.

2. Methods

Aligning with the federated storage approach of the MII, the MeDaX toolbox will be applied locally to enrich and integrate available health care data with data from other sources, including biomedical ontologies [6] and public databases. For the current

[1] Corresponding Author: Judith Wodke, E-mail: judith.wodke@uni-greifswald.de.

prototype, dummy FHIR resources are transferred into a Neo4J KG using BioCypher [8] in strict mode. This first implementation serves as an internal proof of concepts and allows us to start developing analysis and querying tools.

3. Results

The MeDaX KG prototype is work in progress and includes dummy FHIR resources representing data from the MII CDS basic modules [2] plus semi-automatically added ontological [7] and provenance information. In addition, converting different data sources into RDF* format is currently under investigation. Prototype testing is accomplished in cooperation with the DIZ at University Medicine Greifswald and aimed at evaluating our ETL-process, at testing features that can be included into the BioCypher input adapter, and at deciding for a visualisation approach.

4. Outlook

Upon approval of prototype functionality, more data sources will be integrated and routines for data quality and similarity scoring will be added. Also, a feature for data declassification (default: classified) and standardised publication of KG structure will be implemented. Published local KG subsets can be combined into a global public MeDaX KG, providing information about the availability of non-public research. For the first beta release, a graphical user interface for querying the MeDaX KG clinic-internally is planned. In summary, MeDaX will provide a combined data resource to clinicians, scientists, and the interested public. The public MeDaX platform will foster biomedical data reuse by improving findability, accessibility, and interoperability of biomedical data gathered and stored at German hospital clinics and other health care providers.

References

[1] Semler SC, et al. German Medical Informatics Initiative – A National Approach to Integrating Health Data from Patient Care and Medical Research. Methods Inf Med. 2018;57(S 01):e50-6.
[2] Ganslandt T, Boeker M, Lobe M, Prasser F, Schepers J, Semler S, et al. Der Kerndatensatz der ¨ Medizininformatik-Initiative: Ein Schritt zur Sekundarnutzung von Versorgungsdaten auf nationaler ¨ Ebene. In: Forum der Medizin-Dokumentation und Medizin-Informatik. vol. 20; 2018. p. 17.
[3] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE international symposium on computer-based medical systems. IEEE; 2013. p. 326-31.
[4] Finlayson SG, LePendu P, Shah NH. Building the graph of medicine from millions of clinical narratives. Scientific data. 2014;1(1):1-9.
[5] Balaur I, Saqi M, Barat A, Lysenko A, Mazein A, Rawlings CJ, et al. EpiGeNet: a graph database of interdependencies between genetic and epigenetic events in colorectal cancer. Journal of Computational Biology. 2017;24(10):969-80.
[6] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research. 2009;37(suppl 2):W170-3.
[7] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3.
[8] Lobentanzer S, Aloy P, Baumbach J, Bohar B, Danhauser K, Dogan T, et al. Democratising Knowledge Representation with BioCypher. ArXiv. 2022:doi: https://arxiv.org/abs/2212.13543.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230090

A Toolchain for Big Data Analyses in the Intelligent Cognitive Operating Room

Elena KIRSCH [a], Denise JUNGER [a], Christian KÜCHERER [a] and Oliver BURGERT [a,1]
[a] *School of Informatics, Reutlingen University, Reutlingen, Germany*
ORCiD ID: Denise Junger https://orcid.org/0000-0002-7895-3210, Christian Kücherer https://orcid.org/0000-0001-5608-482X, Oliver Burgert https://orcid.org/0000-0001-7118-4730

Abstract. This project aims to evaluate existing big data infrastructures for their applicability in the operating room to support medical staff with context-sensitive systems. Requirements for the system design were generated. The project compares different data mining technologies, interfaces, and software system infrastructures with a focus on their usefulness in the peri-operative setting. The lambda architecture was chosen for the proposed system design, which will provide data for both postoperative analysis and real-time support during surgery.

Keywords. Context-aware operating room, surgical data science, cognitive OR

1. Introduction

Supporting medical staff in the intraoperative area is one of the objectives of computer-assisted surgery. Context-sensitive systems record the current situation in the operating room and can thus provide specific information to support the actors in an operating room in a targeted manner. To provide information, various data must be collected, stored, and managed. There is no common information systems infrastructure available to support such context-aware systems regarding data management and data processing.

The goal of this project is to evaluate existing big data infrastructures regarding their applicability in the operating room. The data management system shall support a situation recognition pipeline [1] and thus enable a context-aware operating room.

2. Methods

We first did literature research on intraoperative context-aware systems to generate the requirements for the system design and interviewed the researchers who built the existing context-aware system at Reutlingen University. Based on the derived requirements, we performed systematic literature research on data mining technologies, interfaces in the medical domain, and software system infrastructures with a focus on their applicability in a peri-operative setting.

[1] Corresponding Author: Oliver Burgert, Faculty of Informatics, Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany; E-mail: oliver.burgert@reutlingen-university.de.

As architectural patterns, we analyzed the lambda and kappa architecture [2] which are dedicated to big data applications. As data storage, we compared data lakes and data warehouse architectures. Data processing frameworks like Map-Reduce [3], Spark, Flink, and others [4] as well as six NoSQL databases were analyzed regarding their applicability.

From those building blocks, a system design proposal was generated.

3. Results

High-level use cases are a) providing data for postoperative analysis (e.g. for surgical data science or for machine learning) and b) providing near-real-time data intra-operatively for context-aware support. Data will be provided via the standards DICOM (imaging), HL7, FHIR (patient data), or SDC (ieee 11073, surgical device data).

Due to the fact that it can be important in the intraoperative area to ensure both real-time data processing and the storage of a master data set, the lambda architecture was chosen. For our prototype, we selected the Apache Hadoop[2] framework, and data storage was realized as a data lake. Figure 1 shows, how surgical device data, e.g. an event generated by the OR light, will be transferred to the data management system.

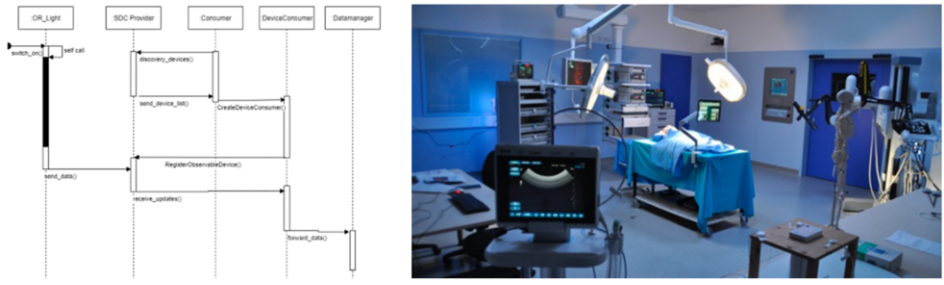


Figure 1. Sequence flow depicting how SDC data is transferred to the data management system (left) and mock operating room testbed for the system at Reutlingen University (right).

4. Discussion and Conclusion

We propose a system for the storage of peri-operative data for context-aware surgical assist systems. The next project step will be the evaluation in a real clinical setting after a successful evaluation in our mock operating room.

References

[1] Junger D, Hirt B, Burgert O (2021) Concept and basic framework prototype for a flexible and intervention-independent situation recognition system in the OR. Comput Methods Biomech Biomed Eng: Imaging Vis 10(3):283-288. https://doi.org/10.1080/21681163.2021.2004446.
[2] Jimmy Lin. The lambda and the kappa. IEEE Internet Computing, 21(5):60–66, 2017
[3] Jeffrey Dean and Sanjay Ghemawat. Mapreduce. Communications of the ACM, 51(1):107–113, 2008.
[4] Jin Wang, Yagiong Yang, Tian Wang, Simon Sherratt, and Jingyu Zhang. Big data service architecture: A survey. Journal of Internet Technology, 2020(21):393–405, 2020.

[2] https://hadoop.apache.org/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230091

Classification of Healthcare Professionals

Annina MAAG [a,1], Renate RANEGGER [a] and Dieter BAUMBERGER [a]
[a] *LEP AG, Rosenbergstrasse 32, 9000 St. Gallen, Switzerland*

Abstract. To be able to compare job titles in healthcare, a proposal for a classification of healthcare professionals was developed. The proposed LEP classification for healthcare professionals is suitable for Switzerland, Germany and Austria and includes nurses, midwives, social workers and other professionals.

Keywords. Occupation, Classification, Healthcare Professionals

1. Introduction

Optimally allocating tasks to healthcare professionals based on their educational background is a key focus of healthcare management [1]. For such analyses, data on the healthcare interventions and on the health care providers are needed [2]. For the measurement of healthcare interventions, LEP [2] classifications can be used; for healthcare professionals, ISCO[3] was found to be lacking in specificity for such analyses in the healthcare sector [3]. In order to be able to analyze more specific data on services and service providers, a proposal is presented for a classification of healthcare professionals.

2. Method

First, a literature review and unstructured interviews with four experts were conducted on the current status of professional education and continuing training in the healthcare sector in Switzerland, Germany and Austria. Subsequently, a model (Figure 1) was created to systematically capture the different levels of education and training.

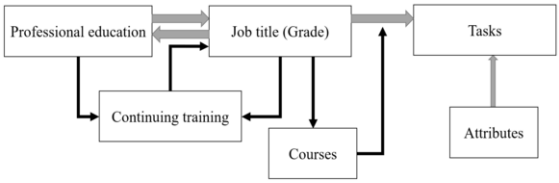


Figure 1. Model of the structure of the healthcare professions for task allocation.

The basis is the professional education, which results in the job title. The job title can be modified or supplemented by further education or training. Function refers to the

[1] Corresponding author: Annina Maag, E-mail: annina.maag@lep.ch
[2] Leistungserfassung in der Pflege, "documentation of nursing activities"
[3] International Standard Classification of Occupations

tasks which are related to the job title and are carried out on an everyday basis. The tasks are dependent on the job title. Courses can influence the function, but they are not relevant for the job title. With the attribution, the function can be further specified. Currently, for example, the level of education or training can be specified, e.g. first semester or onboarding. In a further step, the different levels of education and training were assigned to the occupational titles in healthcare that were researched and a proposal for the classification of occupations was created. Finally, the proposed classification for healthcare professionals was reviewed by two other experts.

3. Results

The proposed LEP classification for healthcare professionals includes nurses, midwives, social workers and the category "other professionals". The LEP classification is structured according to four hierarchical criteria based on levels in a mono-hierarchical structure. The four hierarchical levels are referred to as increasing levels of aggregation. The classification of professions can be used as a master catalogue for international comparisons. Due to the division into sub-catalogues, country specifics such as those pertaining to Switzerland, Germany and Austria can be taken into account (Table 1).

Table 1. Proposed LEP classification for healthcare professionals, with four hierarchical levels, master- and sub-catalogues

Levels	Master	Sub-sub-catalogue (Switzerland)	Sub-sub-catalogue (Austria)
1	Social workers	Social workers	Social workers
2	Higher Vocational Education Master University	Higher Vocational Education Master University	Higher Vocational Education Master University
3	Master of Arts	Master of Arts	Master of Arts
4	Master of Arts in Social Work	Master of Arts in Social Work	Master of Arts in Social Work

4. Discussion and Conclusion

This LEP classification for healthcare professionals, intergraded in the electronic health record and the personnel deployment planning system, could help healthcare managers to establish an overview or the comparability of resource distribution by analyzing services (health care interventions) and service providers (professionals). As a next step, this proposed classification for healthcare professionals must be implemented in a software application (electronic health record, personnel deployment planning system). This would allow it to be tested to show whether the classification can be used in clinical practice and adequately support statistical analyses or whether additional classification elements need to be considered in future developments. The key question is the applicability of this classification in the clinical setting.

References

[1] Drennan, Vari M.; Ross, Fiona (2019): "Global nurse shortages - the facts, the impact and action for change." In: *British Medical Bulletin* 130 (1), pp. 25–37. DOI: 10.1093/bmb/ldz014.

[2] Ranegger R, Baumberger D, Bürgin R. *Demand- and competence-oriented workforce planning according to GuKG 2016 - an activity analysis, (Part A)* [Report]. St. Gallen, Switzerland: LEP AG, 2017. Available at: https://www.lep.ch/de/kontakt.html. Accessed January 06, 2023.

[3] International Standard Classification of Occupations Introduction to occupational classifications (ISCO) https://isco-ilo.netlify.app/en/. Accessed January 06, 2023.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230092

An Enhanced Standardization and Qualification Mechanism for Heterogeneous Healthcare Data

George MANIAS[a,1], Ainhoa AZQUETA-ALZÚAZ[b], Andrea DAMIANI[c], Eshita DHAR[d], Eleftheria KOUREMENOU[a], Marta PATINO-MARTÍNEZ[b], Mariachiara SAVINO[c], Syed Abdul SHABBIR[d] and Dimosthenis KYRIAZIS[a]
[a] *University of Piraeus, Piraeus, Greece*
[b] *Universidad Politécnica de Madrid, Madrid, Spain*
[c] *Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy*
[d] *Taipei Medical University, Taipei, Taiwan*

Abstract. Given the challenge that healthcare related data are being obtained from various sources and in divergent formats there is an emerging need for providing improved and automated techniques and technologies that perform qualification and standardization of these data. The approach presented in this paper introduces a novel mechanism for the cleaning, qualification, and standardization of the collected primary and secondary data types. The latter is realized through the design and implementation of three (3) integrated subcomponents, the Data Cleaner, the Data Qualifier, and the Data Harmonizer that are further evaluated by performing data cleaning, qualification, and harmonization on top of data related to Pancreatic Cancer to further develop enhanced personalized risk assessment and recommendations to individuals.

Keywords. Data Standardization, Data Qualification, Healthcare Analytics

1. Introduction

Nowadays, the healthcare domain faces various challenges related to the diversity and variety of data, their huge volume, and their high distribution. While the massive investments by the healthcare industry into new technologies and the rapid growth in the usage of cloud and mobile computing, medical devices, IoT, and Artificial Intelligence (AI) lead to the increasing need for the design and utilization of enhanced and state-of-the-art healthcare analytics solutions [1]. Thus, the successful cleaning, quality assurance, as well as interpretation and harmonization of the heterogeneous data can provide more precise and personalized prevention and intervention measures, higher experience for patients' health monitoring, and personalized decision support. To address all these challenges this paper introduces a novel mechanism for the cleaning, qualification, and standardization of the collected data. The next section states state-of-the-art standardization and quality assurance techniques, introduces the novel mechanism and its subcomponents, and concludes this research work.

[1] Corresponding Author: George Manias, 80, M. Karaoli & A. Dimitriou St., 18534 Piraeus, Greece; E-mail: gmanias@unipi.gr.

2. Standardization and Qualification Methods and Proposed Approach

Recently, an extensive literature review was performed to existing Clinical Prediction Models (CPMs) [2] examining the different data cleaning approaches that are applied on them. The results indicated that all of them are described by diversity, inconsistency, and lack of reported details in how missing data are handled. As concerns the quality assurance of the healthcare data, different qualitative and quantitative measures and methods are proposed to assess the quality of the data [1]. In addition, the ever-increasing usage of EHRs during the last decades enables more effective and efficient data sharing. However, the use of different clinical content and terminology standards across different healthcare organizations is quite common and thus the lack of a wider standardization for data sharing continues to be a major impediment in achieving true data interoperability in the healthcare domain [3]. Under the scopes of this research work, the proposed mechanism seeks to address the aforementioned lacks and challenges in both the qualification and standardization of healthcare data. To this end, the proposed mechanism exploits three (3) processing phases, the cleaning, the qualification, and the harmonization of the data. These phases are realized through the design and implementation of three (3) integrated and automated subcomponents, i.e., the Data Cleaner, the Data Qualifier, and the Data Harmonizer as depicted in Figure 1.

Figure 1. Standardization and Qualification Mechanism

The proposed mechanism assures the incoming data's accuracy, integrity, and quality, while it also provides a decision whether a connected data source will be considered as reliable or not. On top of this, it enhances the interoperability of data through automated standardization techniques that are applied on the cleaned and qualified data. Through its utilization data are presented in a consistent manner irrespective of the data source. The proposed mechanism will be further applied and evaluated in the context of a novel personalized-healthcare framework as realized in the context of the iHELP project [4].

Acknowledgment

The research leading to the results presented in this paper has received funding from the European Union's funded Project iHELP under grant agreement no 101017441.

References

[1] Cichy C, Rass S. An overview of data quality frameworks. IEEE Access. 2019 Feb 15;7:24634-48.
[2] Tsvetanova A et al. Missing data was handled inconsistently in UK prediction models: a review of method used. Journal of Clinical Epidemiology; 2021; 140: 149-158.
[3] Kubben P, Dumontier M, Dekker A. Fundamentals of clinical data science; 2019.
[4] Manias G, et al. iHELP: Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records. In 2021 IEEE Symposium on Computers and Communications (ISCC); 2021; 1-8; IEEE.

Section 2

Health Information Systems, Patient Records, Visualization Tools and Security & Safety

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230094

Towards Safe Conversational Agents in Healthcare

Kerstin DENECKE[a,1]
[a] *Bern University of Applied Sciences, Bern, Switzerland*
ORCiD ID: Kerstin DENECKE https://orcid.org/0000-0001-6691-396X

Abstract. Conversational agents (CA) are becoming very popular to deliver digital health interventions. These dialog-based systems are interacting with patients using natural language which might lead to misunderstandings and misinterpretations. To avoid patient harm, safety of health CA has to be ensured. This paper raises awareness on safety when developing and distributing health CA. For this purpose, we identify and describe facets of safety and make recommendations for ensuring safety in health CA. We distinguish three facets of safety: 1) system safety, 2) patient safety, and 3) perceived safety. System safety comprises data security and privacy which has to be considered when selecting technologies and developing the health CA. Patient safety is related to risk monitoring and risk management, to adverse events and content accuracy. Perceived safety concerns a user's perception of the level of danger and user`s level of comfort during the use. The latter can be supported when data security is guaranteed and relevant information on the system and its capabilities are provided.

Keywords. Conversational agent, chatbot, safety, risk management, adverse events

1. Introduction

Supposed to be intuitively used, conversational agents (CA) are applied not only in customer service, but also in healthcare. Application areas include the collection of the medical history, supporting self-management or even delivering mental health interventions. CA can cause harm or death in users when badly designed and users rely upon them as authoritative source of information [1]. In particular unconstrained user input could lead to misinterpretations by the CA which in turn could result in wrong advice. Given technological advances, it is becoming increasingly difficult for users to distinguish CA from humans [2]. Created to simulate human behavior, a health CA is supposed to build a bond of trust to the patient and deals with health data. This creates a huge demand of assessing safety of those applications.

However, we can recognize a lack of research on errors and their impact on patient safety when applying CA in healthcare. A review of Abd-Alrazaq et al. on effectiveness and safety of using chatbots to improve mental health demonstrated that current systems are not seriously assessed towards safety [3]. They found two randomized controlled trials (RCT) out of 12 included in their review that reported on safety. However, the developers of those two systems concluded the systems are safe because no harm,

[1] Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Institute for Medical Informatics, Quellgasse 21, 2502 Biel, Switzerland, E-mail: kerstin.denecke@bfh.ch.

distress, adverse events or worsening of the depressive symptoms were reported during the trial that was conducted. We believe that the non-reporting of harm or adverse events is insufficient to be able to conclude about the safety of CA. A more comprehensive analysis is essential. This requires a joint understanding of the facets of safety. Bickmore et al. studied patient and consumer safety risks when using CA (Siri, Alexa, Google Assistant) for gathering medical information [4]. They concluded that "relying on such assistance for actionable medical information represents a safety risk for patients". They found when asking questions that require medical expertise, CA failed more than half of the time and recommended actions could have resulted in harm for the user.

The large research interest in health CA and the missing reporting and assessment of safety aspects related to health CA raises the necessity to define the facets of health CA safety. This paper aims at synthesizing facets of safety in the context of health CA to form a common understanding, to make recommendations for ensuring safety in health CA and shape future research endeavors towards safe health CA.

2. Methods

To identify aspects related to safety of health CA, we were interested in concrete solutions of health CA described in literature. In previous work, we conducted a literature search to identify recent papers published between 2010 and 2022 in which health CA are presented. We searched for relevant scientific papers on PubMed, ACM Digital Library, and IEEE Xplore published between 2010 and 2022 and written in English. To identify appropriate literature, we defined the following search string: *(application OR app OR approach OR implementation) AND (chatbot OR bot OR conversation OR conversational user interface) AND (health OR healthcare)*. Only publications were included that were peer-reviewed conference papers or journal articles of original work. The publication had to present a concrete CA applied in healthcare. We excluded papers not dealing with a concrete healthcare-related CA or only describing the design process, reviews or meta-analyses. The review identified 222 relevant papers. Within these papers, we formed a subset of papers where the full text contained one of the terms *safe* (referring to safety) or *adverse* (referring to adverse event). The resulting 112 documents were manually assessed; they were considered for this paper when they reported on any kind of safety assessment or discussed safety aspects of their solutions. 76 papers contained at least one of the keywords, but used in a different context (e.g., a CA on safer sex or referenced papers with "safe" in the title). They were excluded. 36 papers fulfilled the criteria and were used to extract aspects related to safety were extracted. We aggregated the information, and derived facets of safety and recommendations on how to consider these facets in future health CA developments. The recommendations were derived from the retrieved information and from our experiences on working towards a standard evaluation framework for health CA [5–7].

3. Facets of Safety of Health CA

We can distinguish three facets of safety: 1) system safety, 2) patient safety, and 3) perceived safety. System safety concerns the content that is delivered by the health CA to the user, content reliability, correctness, data quality, data security and data transfer [8–10]. Patient safety concerns the risk that interacting with the health CA could harm a

patient, and perceived safety is the «user's perception of the level of danger and his/her level of comfort during the use" [11]. As follows, we describe the facets and make recommendations to address the safety aspects in the development of CA (Table 1).

In the reviewed papers, system safety primarily concerns data security: Health dialog is characterized by a secure environment. The patient can be sure that her information is only used for the treatment process. This creates a high degree of trust [2]. In case users of a health CA are not confident that data privacy is ensured, they will be less willing to disclose correct information. Health CA must be designed with data privacy in mind, not just for legal requirements but also so that vulnerable patients may develop the confidence necessary to provide personal information.

Table 1. Safety facets with recommendations how to address them during health CA development

Safety facet	Questions to consider	Recommendations
System safety	Is data privacy and data security ensured? Is a data privacy policy available in the health CA?	Robust against cyber attacks Penetration testing Ethical hacking [12] Assessing technical errors (including language understanding and response generation [5]) Providing information on data security and processing including information on use of third part tools
Patient safety	Who is the user? Which safe-critical situations might occur while interacting with the system? Which unexpected consequences might arise (e.g. app-app interactions, adverse events, worsening of symptoms) Who is responsible when a risk occurs? Is only accurate, evidence-based information included in the CA responses and questions? Is the underlying knowledge base evidence-based? Were physicians / healthcare professionals involved in the content development of the health CA? Is there a maintenance process for the information included in the health CA? Is information on the developer or content provider of the health CA provided? Were patient organizations involved in the development of the health CA?	Ensure content accuracy by design Include safety measures for emergencies (e.g. redirect to emergency resources) Clearly describe the limitations of the health CA [13] Assess the adverse events and side effects of the health CA Include automatic risk classifiers (e.g. a self-harm risk classifier [14], or classifier for risks of suicide and violence [15])
Perceived safety	Which personal identifiable information is required and is only this information collected and stored? Does the health CA provides an environment for the user that is safe for disclosing personal information? Did the health CA only provides evidence-based information? Is the information provided in a way that is understandable by the user group of consideration?	Provide information on data use, sharing and storage Provide information on underlying clinical evidence base Provide information on information sources Consider eHealth and health literacy of the user Consider reading level of the user

Two aspects concern patient safety: a) medical safety (does the CA worsen or produce symptoms or diseases in a patient?), and b) emergency safety (Is fast assistance ensured in case of emergencies). In particular when a user is supposed to interact with a health CA for a medium or long-term period, the user might be confronted with a safety-critical situation. When the health CA is autonomous, it has to be ensured that it can recognize such situation and react appropriately [16]. Given the language-based

interaction in the context of health CA, safety problems could result from misunderstandings or misinterpretations of (unconstrained) user input, from inappropriate reaction to unexpected user input, missing knowledge-based interpretation of user input (e.g. drug-drug interactions are not recognized [1]) and missing safety measures when misunderstanding or unexpected user input occurred. To ensure patient safety, mechanisms have to be in place to handle potential health risks such as suicidality, violence or risk of self-harm [17]. This can be realized by including safety plans, generating warnings when a risk is determined, sending appropriate referrals to emergency hotlines, or other contact persons. The content has to be accurate which can be ensured by involving healthcare professionals in the development and relying upon clinical evidence.

Perceived safety can depend on system safety such as data security. Patients reported to feel safe in disclosing information because they have the impression of sharing information with themselves [18]. This is only possible when system safety in general and data security in particular is guaranteed. There are several open issues related to perceived safety: How to measure perceived safety of CA users in healthcare? When do patients feel safe using a CA? What features of health CA are required to feel safe? A user survey could help identifying answers to latter questions.

4. Discussion and Conclusion

In this paper, we described three facets of safety related to health CA: system, patient and perceived safety and we made recommendation how to ensure safety of health CA. There exists a safety event taxonomy, the JCAHO patient safety event taxonomy [19]. This taxonomy describes communication as one of the processes in healthcare that can be faulty or fail and may therefore result in adverse events. It is therefore essential, to study the safety of health CA, since the interaction between a health CA and its users is realized as communication. However, a standardized methodology for assessing safety of health CA along these three facets is still missing. Jang et al. used a questionnaire comprising five aspects to assess Cas` side effects [20]. The aspects are: a disease specific side effect (increase of negative emotional experiences), privacy infringement, sense of alienation from everyday life, violation of therapeutic boundaries, regression of the therapeutic process. Miner et al. studied the CA`s reaction to emergency situations. They analyzed CA`s responses to short emergency messages posted by users [21]. Even though highly relevant, these approaches study only some aspects of the three facets of safety we identified in this work.

The papers reporting on clinical trials with health CA sometimes conclude that the system is “safe to use” since no adverse events were reported by the users [22]. The question arises what is considered as adverse event. Having in mind adverse events in clinical trials, this could be symptoms that occurred during the study period. But when interacting with a machine, additional adverse events may arise, that researchers currently not analyze (e.g., app-app interactions, upcoming addictions to the technology, impact on social activities). To address these issues, we recommend future research on possible adverse events of health CA. This could result in a taxonomy of such events and safety aspects which would contribute to a common view of relevant and possible adverse events. A harmonized safety risks assessment framework could help in improving the assessment of safety risks due to health CA usage. Finally, developing a reporting guideline for safety assessment in health CA would support transparency. From

a patient perspective, an information sheet on possible adverse events and safety risks similar to the package inlet for drugs could help in increase perceived safety and would contribute to an informed patient who is reflecting critically the symptoms that occur and can ask for professional help when necessary.

References

[1] Bickmore T, Trinh H, Asadi R, Olafsson S. Safety First: Conversational Agents for Health Care. In: Moore RJ, Szymanski MH, Arar R, Ren G-J, editors. Studies in Conversational UX Design [Internet]. Cham: Springer International Publishing; 2018 [cited 2022 Nov 6]. p. 33–57. Available from: http://link.springer.com/10.1007/978-3-319-95579-7_3

[2] Bickmore T, Giorgino T. Health dialog systems for patients and consumers. Journal of Biomedical Informatics. 2006;39:556–71.

[3] Abd-Alrazaq AA, et al. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res. 2020;22:e16021.

[4] Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res. 2018;20:e11510.

[5] Denecke K, Abd-Alrazaq A, Househ M, Warren J. Evaluation Metrics for Health Chatbots: A Delphi Study. Methods Inf Med. 2021;60:171–9.

[6] May R, Denecke K. Security, privacy, and healthcare-related conversational agents: a scoping review. Informatics for Health and Social Care. 2022;47:194–210.

[7] Denecke K, May R. Usability Assessment of Conversational Agents in Healthcare: A Literature Review. In: Séroussi B, Weber P, Dhombres F, Grouin C, Liebe J-D, Pelayo S, et al., editors. Studies in Health Technology and Informatics [Internet]. IOS Press; 2022 [cited 2022 Jun 20].

[8] Valtolina S, Hu L. Charlie: A chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter [Internet]. Bolzano Italy: ACM; 2021 [cited 2022 Jan 13]. p. 1–5.

[9] Tschanz M, et al. Using eMMA to Manage Medication. Computer. 2018;51:18–25.

[10] Ollier J, et al. Elena+ Care for COVID-19, a Pandemic Lifestyle Care Intervention: Intervention Design and Study Protocol. Front Public Health. 2021;9:625640.

[11] Amini R, Lisetti C, Yasavur U, Rishe N. On-Demand Virtual Health Counselor for Delivering Behavior-Change Health Interventions. 2013 IEEE International Conference on Healthcare Informatics. 2013. p. 46–55.

[12] Ganapathy S, et al. Acute paediatrics tele-support for caregivers in Singapore: an initial experience with a prototype Chatbot: UPAL. Singapore Med J. Singapore; 2021;

[13] Mauriello ML, et al. A Suite of Mobile Conversational Agents for Daily Stress Management (Popbots): Mixed Methods Exploratory Study. JMIR Form Res. 2021;5:e25294.

[14] Deshpande S, Warren J. Self-Harm Detection for Mental Health Chatbots. Stud Health Technol Inform. Netherlands; 2021;281:48–52.

[15] Collins C, et al. Covid Connect: Chat-Driven Anonymous Story-Sharing for Peer Support. Designing Interactive Systems Conference. Virtual Event Australia: ACM; 2022. p. 301–18. 5

[16] Tielman M, et al. A Therapy System for Post-Traumatic Stress Disorder Using a Virtual Agent and Virtual Storytelling to Reconstruct Traumatic Memories. Journal of medical systems. 2017;41:125.

[17] Hungerbuehler I, et al. Chatbot-Based Assessment of Employees' Mental Health: Design Process and Pilot Implementation. JMIR Form Res. 2021;5:e21678.

[18] Brandtzaeg PB, Følstad A. Chatbots: Changing User Needs and Motivations. Interactions. New York, NY, USA: Association for Computing Machinery; 2018;25:38–43.

[19] Chang A, Schyve PM, Croteau RJ, O'Leary DS, Loeb JM. The JCAHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events. Int J Qual Health Care. 2005;17:95–105.

[20] Jang S, Kim J-J, Kim S-J, Hong J, Kim S, Kim E. Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit: A development and feasibility/usability study. Int J Med Inform. Ireland; 2021;150:104440.

[21] Miner AS, et al. Talking to Machines About Personal Mental Health Problems. JAMA. 2017;318:1217.

[22] Maher CA, et al. A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study. JMIR Mhealth Uhealth. 2020;8:e17558.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230095

Development of a Command Line Interface for the Analysis of Result Sets from Automated Queries to Literature Databases

Tobias J. BRIX[a,1] Malte BERENTZEN[b], Ludger BECKER[b], Michael STORCK[a] and Julian VARGHESE[a]
[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Institute of Computer Science, University of Münster, Germany*
ORCiD ID: Tobias J. Brix https://orcid.org/0000-0002-1557-8097

Abstract. The first step of a systematic review is the identification of publications related to a research question in different literature databases. The quality of the final review is mainly influenced by finding the best search query resulting in high precision and recall. Usually, this process is iterative and requires refining the initial query and comparing the different result sets. Furthermore, result sets of different literature databases must be compared as well. Objective of this work is to develop a command line interface, which supports the automated comparison of result sets of publications from literature databases. The tool should incorporate existing application programming interfaces of literature database and should be integrable into more complex analysis scripts. We present a command line interface written in Python and available as open-source application at https://imigitlab.uni-muenster.de/published/literature-cli under MIT license. The tool calculates the intersection and differences of the result sets of multiple queries on a single literature database or of the same query on different databases. These results and their configurable metadata can be exported as CSV-files or in Research Information System format for post-processing or as starting point for a systematic review. Due to the support of inline parameters, the tool can be integrated into existing analysis scripts. Currently, the literature databases PubMed and DBLP are supported, but the tool can easily be extended to support any literature database providing a web-based application programming interface.

Keywords. Systematic Review, Literature Database, PubMed, DBLP

1. Introduction

Systematic reviews are important for the scientific community to provide an overview of the current state-of-the-art of a given topic or an entire research field. Guidelines for performing systematic reviews have been proposed to ensure reproducibility. At the time of writing, most journals request the authors to follow the PRISMA guidelines [1]. As first step of this guideline, a search string must be defined, which is used to query scientific articles from multiple literature databases. During the development of the search string, search terms are often exchanged, e.g., by synonyms, or different search

[1] Corresponding Author: Tobias J. Brix, Institute of Medical Informatics, University of Münster, Albert-Schweitzer-Campus 1, Building A11, 48149 Münster, Germany; E-mail: tobias.brix@uni-muenster.de

terms are connected via logical operators like “and” or “or”. After each iteration with different numbers of results, the main task is to identify the newly found articles and articles, which were missed by the changed query. Mathematically speaking, having two article result sets from different queries, A and B, the set difference A\B and B\A must be calculated and analyzed. When the final search string has been determined, it is used to query different literature databases and duplicates must be filtered out, i.e., the intersection of two result sets A∩B must be determined. These calculations involving thousands of articles are very time consuming and error prone. Even though most citation managers support the filtering of duplicates, the export of result sets and import into the manager is not feasible for rapid prototyping of queries.

Objective of this work is to provide a command line interface supporting the operations described above on result sets from search queries on literature databases. The tool should be capable of sending queries to literature databases automatically via their web-based application programming interfaces (APIs) and support the integration into existing analysis scripts via inline parameters. The tool should be flexible enough to allow the extension to any literature database providing an API.

2. Methods

Literature or academic databases are large collections of references to peer-reviewed scientific research articles. Besides the references themselves, these databases store structured metadata, like authors’ names or journal information, to support detailed search queries and allow scientifically valid citations. Most literature databases provide their own online search engine on these metadata and an API for querying articles programmatically. Usually, these literature databases are focused on specific fields of research. In this work, we will exemplarily focus on PubMed (www.pubmed.gov) and DBLP (www.dblp.org).

PubMed is a free to use literature database covering life sciences and biomedical topics [2]. In 1997, the database was first available to the public and is currently maintained by the United States National Library of Medicine (NLM) at the National Institutes of Health (NIH). At the time of writing, PubMed consists of over 34 million scientific references. Its search engine supports the search for phrases, i.e., exact order of words, and any combination of logical operations “and”, “or” and “not”, grouped by brackets. In addition, search terms can be limited to a certain metadata filed, e.g., author’s name, by using a postfix notation. The search API is well documented and free to use.

The Digital Bibliography & Library Project (DBLP) is a literature database containing conference and journal articles from the field of computer science [3]. DBLP was founded in 1993 by Michael Ley at the University of Trier. Since 2018, the database is hosted and maintained as a service by the Leibniz Center for Informatics Schloss Dagstuhl. At the time of writing, DBLP consists of over 6 million references. Its search engine does not support the search for phrases, however, by using “-” like “first-second” it can be enforced that “second” must appear anywhere after “first”. By default, each word of the search string is implicitly connected via an “and”. The logical “or” is only supported on the level of single words. The logical “not” as well as groupings and brackets is not supported and will be ignored. Similar to PubMed, each search term can be limited to a certain metadata field by using a prefix notation. Again, the search API is well documented and free to use but limited to 10000 references per query result.

3. Results

3.1. The Command Line Interface

The application is available as open-source software under the MIT license. It is written in Python 3.11 and provides a command line interface for the programmatic integration into more complex analysis scripts. The tool can be called with "py literature-cli.py <parameters>". The parameter "-h" shows the instructions and a list of available parameters. The tool offers two modes, which support sending multiple queries to a single literature database or a single query to multiple databases. For the remainder of this section, we will focus on the second use case since both modes work similarly.

First, the query is transformed into the syntax of each supported database. Logical operations like "AND", "OR" and "NOT" are replaced according to the requirements of each literature database, e.g., in case of DBLP the replacement of "OR" with "|" and removal of all "AND" operators, which are implicit.

Afterwards, the transformed search string is sent to the corresponding API endpoints. By default, only the metadata of the result set, containing the authors' names, title, journal, publication year and DOI, are further processed. If a DOI is present, it is used to identify a publication in different result sets. Otherwise, the heuristic of matching title and year could be applied since both values are most consistent between different literature databases and are not affected by abbreviations. All combinations of intersections and differences of the result sets are calculated, which, for q queries sent, are $2^q - 1$ sets. These subsets can be exported as separate CSV-files containing all metadata for further post-processing or in the Research Information System (RIS) format for the upload into a citation manager as starting point of a systematic review. Furthermore, if defined in the parameter list, a Venn diagram is provided as visual feedback [4]. To provide a history of all requests and enable reproducibility, all query parameters and results are stored in a logging file as shown in Table 1.

3.2. Configuration and Extensibility

The internal structure of the tool is designed to enable easy extensibility to additional databases, as long as they provide a web-based API. For each new database, a class must be implemented inheriting from the connection base interface. The class implements the specific API endpoints, because structure and call-order are highly literature database dependent. Besides the implementation, a configuration file must be provided, which contains the exact location of metadata in the response. The associated XML-tags in the response can be defined by XPath. Furthermore, a mapping to the tool's syntax must be provided, e.g., the handling of logical operations like "AND", "OR" and "NOT".

The metadata, which identifies a publication, if no DOI is provided, can be configured application-wide. In addition, all metadata, which should be exported or taken into consideration during analysis can be specified and extended as well. These changes must be replicated to the individual configuration of each connected database, i.e., the path to the corresponding XML element in the response message.

3.3. Example Workflow

The tool is most useful when processing large quantities of results that can hardly be managed manually. For comprehensibility, a minimal example is considered here. Let us

assume that a systematic review about machine learning approaches in the context of the rare Kawasaki disease should be performed [5]. Since it is an interdisciplinary topic between computer science and medicine, the literature databases DBLP and PubMed should be included in the systematic review. In the following, the process of finding the appropriate search string using the aforementioned tooling will be explained.

First, we directly experiment on the PubMed website by searching "Kawasaki disease" AND "machine learning" getting 16 results. Then, we try a word with a related meaning and replace "machine" by "deep" and receive 7 results. As shown in line 1 of Table 1, we use the tool to compare both results. There is only an overlap of 2 articles so combining both terms is beneficial. Back on the PubMed website, we combine both terms with the logical "OR" operator and get 21 results as expected. We are satisfied with the current result and want to apply it to DBLP. Unfortunately, DBLP does not support the phrase search and only supports OR for single tokens. Therefore, we restructure the query in a way that it can be interpreted correctly by DBLP. As shown in line 2, we verify by using the tool that our transformed query does not lose previous publications in PubMed. Instead, we are now getting 23 results, but keeping all previous 21 publications. Finally, we apply the query to both literature databases and get our result of 26 unique publications for our systematic review as shown in line 3. The corresponding RIS-files can directly be loaded into our citation manager.

Table 1. Logging example of the command line interface. All used parameters and results are documented to track the progress and support reproducibility.

Line	Input	Result
1	py literature-cli.py –m alternate –db pubmed –q ‘ “Kawasaki disease” AND “machine learning” ’ ‘ “Kawasaki disease” AND “deep learning” ‘	q1: 16 q2: 7 q1∩q2: 2
2	py literature-cli.py –m alternate –db pubmed –q ‘ “Kawasaki disease” AND (“machine learning” OR “deep learning”) ‘ ‘ “Kawasaki-disease” AND (machine OR deep) AND learning ‘	q1: 21 q2: 23 q1∩q2: 21
3	py literature-cli.py –v –o ris –q ‘ “Kawasaki-disease” AND (machine OR deep) AND learning ‘	pubmed: 23 dblp: 4 pubmed∩dblp: 1

4. Discussion

All pre-defined requirements of the tool have been met. As illustrated by the example workflow, it can help identifying publications during a systematic review. By providing it as open-source, it can be used by other researchers freely.

Tools like DistillerSR (www.distillersr.com) or Rayyan [6] also support systematic reviews. However, both are not open-source. They use advanced artificial intelligence to filter duplicate references and guide the user through the entire PRISMA workflow. In case of DistillerSR, even an API call to PubMed is supported. Nevertheless, these tools require a final query or already exported lists of references. Thus, our tool can be applied as a kind of preprocessing to determine the required references in a rapid-prototyping fashion, before the main PRISMA workflow begins.

A few limitations need to be addressed though. The tool was designed as a command line interface to integrate it programmatically into complex analysis scripts. The nature of the command line may be off-putting to technically unsophisticated users or even limit its usability. A graphical user interface could certainly promote acceptance in the context

of future work. Until then, there is still the possibility to use the search engines of the literature databases directly as shown in the example workflow.

Secondly, only two literature databases are currently integrated, primarily due to problems with not freely usable APIs. Well-known literature databases such as Web of Science (www.webofscience.com) or Scopus (www.scopus.com) have paid licenses or usage restrictions (number of hits per week), which make meaningful free use harder. Another example is the literature database Google Scholar (scholar.google.de), which requires a paid third party provider license (www.serpapi.com).

The support of search tags and a general pagination approach are planned as future work. Many literature databases allow the restriction of individual search terms to specific metadata, such as searching only for authors' names, e.g., by using the postfix "[AU]" in PubMed or prefix "author:" in DBLP. The usage of such terms is generally supported by the application. However, corresponding queries should only be sent to a single database, since the query can be misinterpreted by other databases due to the greatly varying syntax. This issue could be handled similarly to logical operators during query conversion by applying mappings from the configuration files.

Some APIs limit the number of results returned per request. Currently, the maximum number of possible results of a single query is delivered, which is 10000 hits for PubMed and DBLP. Results above 10000 hits are considered as too low precision and are therefore ignored. In the future, this should be addressed by dynamically loading all publications, if pagination is supported by the API or heuristically implemented by adding the publication year to the search string and iterating over each year that contains at least one publication in the result set.

5. Conclusions

In this work, we presented a command line tool, for the calculation of intersection and differences of article result sets from automated queries to literature databases. The tool fulfills all pre-defined requirements and can help during the process of conducting a systematic review. Currently, the connection to PubMed and DBLP is implemented but a connection to further databases providing an API can easily be added. The source code is available from https://imigitlab.uni-muenster.de/published/literature-cli.

6. References

[1] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ (Clinical research ed.) 2021;372:n71.

[2] Lindberg DA. Internet access to the National Library of Medicine. Effective clinical practice ECP 2000;3(5):256–60.

[3] Ley M. The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives: In: Goos G, Hartmanis J, van Leeuwen J, Laender AHF, Oliveira AL, editors. String Processing and Information Retrieval. Volume 2476. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002, p. 1–10.

[4] Venn J. I. On the diagrammatic and mechanical representation of propositions and reasonings. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 1880;10(59):1–18.

[5] Gorelik M, Chung SA, Ardalan K, Binstadt BA, Friedman K, Hayward K et al. 2021 American College of Rheumatology/Vasculitis Foundation Guideline for the Management of Kawasaki Disease. Arthritis care & research 2022;74(4):538–48.

[6] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Systematic reviews 2016;5(1):210.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230096

Cross-Registry Benchmarking of Data Quality: Lessons Learned

Jürgen STAUSBERG[a,1], Sonja HARKENER[a], Christoph ENGEL[b], Robert FINGER[c], Carsten HEINZ[d], Ekkehart JENETZKY[e,f], Patrick JERSCH[g], David MARTIN[e,h], Rüdiger RUPP[g], Martin SCHOENTHALER[i], Barbara SUWELACK[j], and Jeannine WEGNER[j]

[a] *University Duisburg-Essen, Faculty of Medicine, IMIBE, Essen, Germany*
[b] *Leipzig University, IMISE, Leipzig, Germany*
[c] *Department of Ophthalmology, University Hospital Bonn, Bonn, Germany*
[d] *Department of Ophthalmology, St. Franziskus-Hospital Münster, Münster, Germany*
[e] *Faculty of Health/School of Medicine, Witten/Herdecke University, Witten, Germany*
[f] *University Medical Center of the Johannes-Gutenberg-University, Mainz, Germany*
[h] *Spinal Cord Injury Center, Heidelberg University Hospital, Heidelberg, Germany*
[g] *Department of Pediatrics, Eberhard-Karls University Tübingen, Tübingen, Germany*
[i] *Department of Urology, University of Freiburg, Freiburg, Germany*
[j] *University Hospital Münster, Transplant Nephrology, Münster, Germany*
ORCiD ID: Jürgen Stausberg https://orcid.org/0000-0003-0348-5579

Abstract. Feedback of data quality measures to study sites is an established procedure in the management of registries. Comparisons of data quality between registries as a whole are missing. We implemented a cross-registry benchmarking of data quality within the field of health services research for six projects. Five (2020) and six (2021) quality indicators were selected from a national recommendation. The calculation of the indicators was adjusted to the registries' specific settings. Nineteen (2020) and 29 results (2021) could be included in the yearly quality report. Seventy-four per cent (2020) and 79% (2021) of the results did not include the threshold in their 95%-confidence-limits. The benchmarking revealed several starting points for a weak-point analysis through a comparison of results with a predefined threshold as well as through comparisons among each other. In the future, a cross-registry benchmarking might be part of services provided through a health services research infrastructure.

Keywords. Benchmarking, data quality, quality indicator, registry

1. Introduction

Medical registries provide an understanding about daily health care based upon an observational recording of health-related information. The understanding of daily health care is a prerequisite for advanced tasks as quality research, health economics, or benefit assessment. Having recorded data as its assets, the quality of the data determines the level of knowledge a registry is able to offer. Data should allow, on the one hand, the

[1] Corresponding Author: Prof. Dr. med. Jürgen Stausberg, IMIBE, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; E-mail: stausberg@ekmed.de.

answering of research questions defined in a registry protocol and, on the other hand, the execution of statistical procedures defined in an analysis plan. The data itself could be checked concerning inherent characteristics such as data completeness, consistency, or timeliness. Typically, a data management and monitoring facility is responsible for the quality control of the data. Recommendations about the organization and the operation of such a facility are published [1]. One element of quality control could be a feedback to study sites about the quality of the data for which a study site is responsible. Feedback is a well-known intervention of quality management in health care [2]. Current results of performance measures of one study site could be compared with results of other sites as well as with predefined target values. Missing is the possibility to compare the level of data quality of one registry with the level of data quality of another registry. This cross-registry benchmarking would allow learning from the best in the field, not only for study sites within a registry but also for the registry as a whole.

A funding initiative of the German Ministry of Education and Research brought together six medical registries and a supporting accompanying project. The registries cover different medical fields and objectives: ParaReg documents inpatient stays of people with spinal cord injury or disorder, families record fever episodes for the FeverApp-registry, TOFU is interested in treatment exit options for non-infectious, non-anterior uveitis patients with the lowest rate of recurrence, HerediCaRe interconnects health care and health research for women at risk of ovarian and breast cancer, RECUR integrates data of patients suffering from recurrent calculus of the urinary tract from a smartphone app with data from university clinics, living donors of kidney transplants are recruited by SOLKID-GNR (cf. [3] for details about the registries). The six projects started in 2019 with the realization of their registries. In parallel, the accompanying project implemented the cross-registry benchmarking in close cooperation with the registries. In this paper, we want to share our experiences and lessons learned so far.

2. Material and Methods

2.1. Selection of quality indicators

The set of quality indicators used for the cross-registry benchmarking had to be based on an available national recommendation [4]. Two approaches were combined to reduce this set of 51 indicators to a feasible volume. On the one hand, 15 projects in a preceding phase of the funding initiative rated the indicators [5]. On the other hand, an estimation about the relevance of the quality indicators was adopted from the literature [6]. This selection led to an initial set of five indicators (keys in parenthesis): missing values in mandatory data elements (TMF-1014), outliers (TMF-1018), recruitment rate (TMF-1030), drop-out-rate (TMF-1034), completeness (TMF-1046). After the first quality report, two indicators were excluded. TMF-1018 was found to be not helpful due to optimal results for all registries. The application of TMF-1046 was too complex for the moment [7]. Therefore, the subsequent version of the set still included TMF-1014, TMF-1030 and TMF-1034. Three other indicators were added: consistency (TMF-1003), observational units with follow-up (TMF-1042), residual classes for qualitative data elements (REGISVF-1053). Most of the results were rates, having the enumerator as part of the denominator. The calculations of TMF-1034 and TMF-1042 led to ratios with a denominator population of a preceding period. TMF-1030 was a rate in case of registries aiming for case completeness, a ratio in case of registries with fixed target size.

2.2. Organization of the benchmarking

The available definitions of the quality indicators [4] were extended for the concrete use case. In order to achieve fair comparisons, an adjustment to the different settings was added. Following the definition of data quality in ISO 8000 [8], the adjustment could A) vary the applied quality indicators between the registries, B) define different requirements for each registry, or C) use registry-specific thresholds. We decided to apply a common set of quality indicators with only one general threshold for each indicator in the benchmarking. Adjustment was implemented by offering the registries option B, to adapt the calculation of the quality indicators to their conditions. For example, each registry determined the data elements included in the calculation of TMF-1014, the number and the content of integrity rules checked for TMF-1003, or the denominator used in TMF-1030. Threshold values for the indicators were taken from the literature [4] where possible.

Due to data protection reasons, the registries themselves calculated the results based on the specification of the quality indicators and a joint compilation of the registry-specific adaptations. The accompanying project received the results from each registry in a filled template with raw figures and rates or ratios at the beginning of the subsequent year. The accompanying project supplemented the results with 95%-confidence-intervals (95%-CI). A quality report covering all registries with the selected quality indicators was created including tabular and graphical elements. To support the registries in the interpretation of their results, remarks indicated noticeable results that could be the starting point of a weak point analysis. The quality reports were discussed together in workshops and additionally in bilateral meetings.

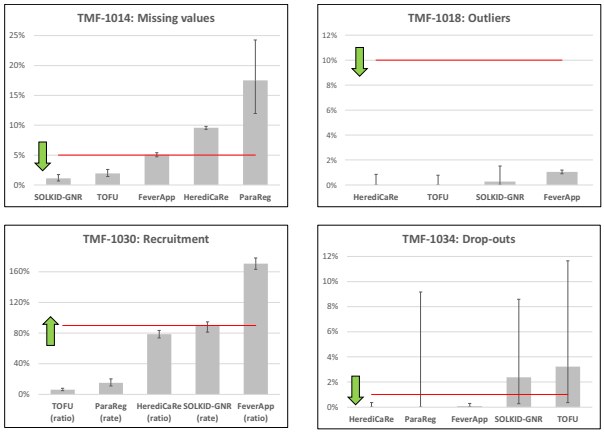


Figure 1. Results of the data year 2020. Arrows indicate the desired direction, the horizontal lines the predefined threshold, and the vertical lines the 95%-CI.

3. Results

Quality reports were created for the data years 2020 and 2021. Due to the Corona pandemic, recruitment of patients started delayed and slower as expected. One registry (RECUR) was not able to recruit any patient in these years. ParaReg did not record quantitative data elements with the possibility of outliers in 2020, in 2021 there was not

any case with expected follow-up. Completeness (TMF-1046) was excluded from the quality report (cf. [7] for a discussion of the reasons). We ended up with 19 values for four quality indicators in the report 2020 (cf. figure 1) and 29 results for six quality indicators in the report 2021 (cf. figure 2).

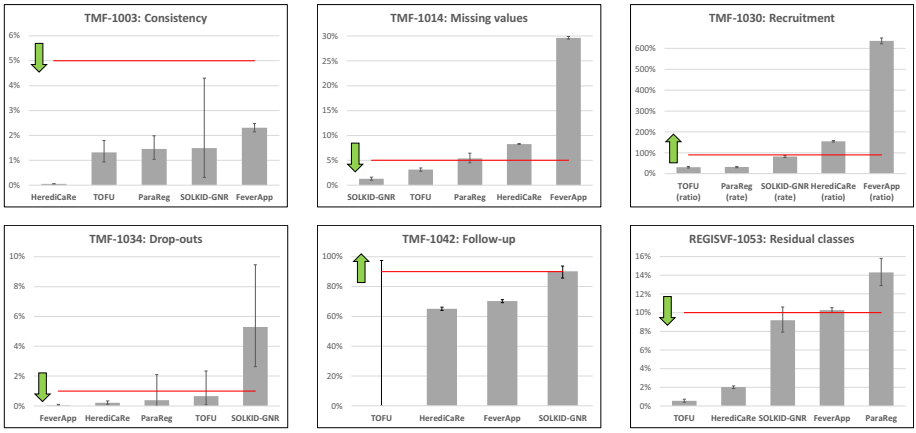


Figure 2. Results of the data year 2021. Arrows indicate the desired direction, the horizontal lines the predefined threshold, and the vertical lines the 95%-CI.

4. Discussion

We expected that the cross-registry benchmarking of data quality could be implemented like a cherry on top of carefully designed case report forms, a systematic and well-founded central monitoring, regularly scheduled reports for the registry office, audits at study sites with at least a partial source data verification, and a benchmarking of study sites using a wide range of quality indicators. However, we recognized that the basis for the cross-registry benchmarking was very heterogeneous. For example, some indicators were specifically implemented for the benchmarking and were not a by-product of an already implemented monitoring. For some projects, the cross-registry benchmarking produced pressure from the cherry on top, leading to extra efforts. Therefore, we assume that not all positive effects of the benchmarking are visible in the results of the quality indicators. The quality reports also highlighted a reverse correlation between data (TMF-1014) and case (TMF-1030) completeness. Both registries, that achieved outstanding results for the recruitment rate in 2021, had the highest rates of missing values.

Quality indicators should provide reliable and valid results. Good quality indicators support responsible parties in quality improvement [9]. In this respect, the quality reports of the cross-registry benchmarking provided a couple of starting points for further actions. All registries received a critical feedback in view of the thresholds differentiating between good and poor data quality. Additionally, the large span of results for most of the indicators allowed a head-to-head comparison of the registries, initiating a learning experience. Further quality-related actions are the rationale of using indicators for quality management to "identify events that merit further review" [9]. That does not mean that every critical result really illustrates a problem. We feel that a set of indicators should be tailored to the current needs. For example, using electronic systems for data capture in all registries avoided outliers through plausibility checks at the time of data recording.

The respective indicator did not provide any new insights and was skipped. The registries' repository of metadata (cf. https://www.toolpool-gesundheitsforschung.de/produkte/metadaten-bmbf-foerdermassnahme-modellhafte-register) disclosed the frequent use of residual classes as "other" for categorical data elements. A respective, newly added quality indicator was found to be helpful and motivated some actions immediately after sampling the data for the quality report.

5. Conclusions

Due to their heterogeneity, a quality indicator that makes sense in one registry may be nonsensical in another. Therefore it is very important to consider the quality indicators in their respective context. Our experience has also shown that the process of discussing quality indicators between those responsible for registry design and analysis can contribute to improving the respective registries. Being aware of the fact, that implementing a cross-registry benchmarking does not replace a thorough evaluation of the method, we are convinced that this approach is worthwhile to be further investigated. There might be a future with services that allow independent registries to compare their data quality with each other beyond the frame of a funding initiative.

Acknowledgements

Funded by the German Federal Ministry of Education and Research under contracts 01GY1901, 01GY1902, 01GY1903, 01GY1904, 01GY1905, 01GY1906, 01GY1917B.

References

[1] Gliklich RE, Leavy MB, Dreyer NA (sr eds). Registries for Evaluating Patient Outcomes: A User's Guide. 4th ed. AHRQ Publication No. 19(20)-EHC020. Rockville, MD: Agency for Healthcare Research and Quality, September 2020.

[2] Barnett GO, Winickoff R, Dorsey JL, Morgan MM, Lurie RS. Quality Assurance through Automated Monitoring and Concurrent Feedback Using a Computer-Based Medical Information System. Med Care 1978;11:962-70.

[3] Stausberg J, Harkener S, Semler S. Recent trends in patient registries for health services research. Methods of Information in Medicine 2021;60(S 01):e1-e8.

[4] Nonnemacher M, Nasseh D, Stausberg J. Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 2014.

[5] Harkener S, Stausberg J, Hagel C, Siddiqui R. Towards a core set of indicators for data quality of registries. Stud Health Technol Inform. 2019;267:39-45.

[6] Schmidt CO, Krabbe CEM, Schössow J, et al. Quality standards for epidemiologic cohort studies: An evaluated catalogue of requirements for the conduct and preparation of cohort studies. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2018; 61: 65-77 (in German).

[7] Stausberg J, Harkener S, Jenetzky E, Jersch P, Martin D, Rupp R, Schönthaler M. FAIR and Quality Assured Data - The Use Case of Trueness. Stud Health Technol Inform 2022;289:25-8.

[8] ISO 8000-2:2020(en) Data quality - Part 2: Vocabulary (s. https://www.iso.org/obp/ui/#iso:std:iso:8000:-2:ed-4:v1:en).

[9] Joint Commission on Accreditation of Healthcare Organisations (JCAHO). Primer on Indicator Development and Application. Measuring Quality in Health Care. JCAHO: Oakbrook Terrace, 1990.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230097

'Quer N0 AVC' for Monitoring Stroke Patients' Healthcare Using a Mobile App

Gabrielle dos Santos LEANDRO[a,b,1], Ivonei BITTENCOURT[c], Vivian NAGEL[c], Carla Heloisa Cabral MORO[c] and Claudia MORO[a]
[a] *Pontifícia Universidade Católica do Paraná, Curitiba, Brazil*
[b] *Health Education and Innovation Center, Joinville, Brazil*
[c] *Joinville Stroke Registry, Joinville, Brazil*
ORCiD ID: Gabrielle S. Leandro https://orcid.org/0000-0003-0751-3581, Ivonei Bittencourt https://orcid.org/0000-0001-7555-9800, Vivian Nagel https://orcid.org/0000-0002-3321-6110, Carla H.C. Moro https://orcid.org/0000-0001-6346-939X, Claudia Moro https://orcid.org/0000-0003-2637-3086

Abstract. Stroke is one of the leading causes of death and impairments worldwide. After hospital discharge, it is necessary to monitor these patients during their recovery. This research addresses the implementation of a mobile app, entitled 'Quer N0 AVC', to improve the quality of stroke patient care in Joinville, Brazil. The study method was divided into two parts. The adaptation phase included all the necessary information in the app for monitoring stroke patients. The implementation phase aimed to prepare a routine for the Quer mobile app installation. One of the questionnaires collected data from 42 patients and identified that before hospital admission 29% of them did not have medical appointments, 36% had one or two appointments, 11% had three appointments, and 24% had four or more appointments. This research portrayed adaptation feasibility and the implementation of a cell phone app for following up on stroke patients.

Keywords. Stroke, Mobile App, Patient Monitoring.

1. Introduction

Stroke is the second-leading cause of death and the third-leading cause of disability-adjusted life years lost worldwide [1]. After a stroke, individuals continue with a broad range of impairments, such as weakness or paralysis, sensory losses, immobility, spasticity, and stroke-related pain [2]. The stroke patient is cared for in several healthcare institutions during the recovery phase after discharge, such as hospitals, primary healthcare units, outpatient facilities, etc. Due to communication challenges among service providers, an increased risk of care gaps accompanies those changes in institutions during treatment and rehabilitation [3]. Therefore, it is essential to monitor the stroke patients' journey during their rehabilitation [4,5].

The department that currently follows up on stroke patients in Joinville - Brazil is Joinvasc (Joinville Stroke Epidemiological Registry). Joinvasc was constituted by

[1] Corresponding Author: Gabrielle dos Santos Leandro, Graduate Program in Health Technology (PPGTS), Pontifícia Universidade Católica do Paraná (PUCPR), 1155 Imaculada Conceição St., Curitiba, PR, 80215-901, Brazil. E-mail: gabrielle.leandro@pucpr.edu.br

Municipal Law # 7,448 on June 12, 2013. Its purpose is to register data on stroke occurrences in public and private institutions in Joinville, specifying those occurrences epidemiologically to contribute to public health policies on the scopes of stroke prevention, treatment, and patient follow-up [6]. Currently, the Joinvasc employees get in touch with patients 30 and 90 days after being discharged from the hospital to follow up on stroke patients, and annually from 1 year up to 5 years after the stroke occurs. They do that by calling them and asking predefined questions.

In the last decade, mobile healthcare technologies (mHealth) were launched as a promising pathway for improving physician-patient communication. Some studies address the utilization and practicality of mHealth in handling chronic diseases [7]. mHealth provides an opportunity for obtaining and clustering different types of data from several sources through technologies such as application programming interfaces [8]. Through mHealth, patients can input data in apps making their health monitoring feasible through questionnaires incorporating patient-reported outcome measurement data [9,10].

The 'Quer' mobile app is a software program developed by the Univision[2] company, and it is a support tool for healthcare. It is possible to customize clinical information and keep it related; it is explicitly collected based on health context. It is also prepared to participate actively in its healthcare monitoring. Due to its characteristics, Quer mobile app can be adapted to monitor stroke patients and aid in avoiding secondary strokes.

Therefore, the objective of this study was to implement the Quer app for following up on stroke patients. The importance of this case study is that Quer mobile app can partially replace the mechanical process performed through telephone calls. In contrast, the patient (or caregiver) answers various risk-control questions, health maintenance, patient rehabilitation, and updates on their current health status. Thus, using the app may automate and streamline data collection for inputting in the Joinvasc database.

2. Methods

The applied research was exploratory objectives [11,12], including adapting and implementing a mobile app in clinical practice. This study was approved by the Research and Ethics Committee Report # 4,917,962. It was performed at São José Municipal Hospital, in Joinville - Brazil, in the Stroke Unit and Joinvasc. The data collected in this step of the research was prospective. The subjects studied in the first step are patients who suffered a stroke from July 2021 to August 2022, totaling 218 patients. First, it is important to emphasize that before the adaptation and implementation of the technology in clinical settings, a partnership was agreed to with the following institutions: Brazilian Stroke Association, Pontifícia Universidade Católica do Paraná, Univision, Joinville City Hall, and Joinvasc.

This study englobed two main phases. The first was the adaptation phase; in this phase, we used the predefined questions previously used by Joinvasc to draft the questionnaires sent to stroke patients. The adaptation phase included all necessary information already used for monitoring patients who have suffered a stroke episode currently used by Joinvasc. The second phase was the implementation phase, which aimed to implement the use of the Quer mobile app during stroke patient hospitalization and their monitoring after hospital discharge.

[2] Univision website: https://univision.net.br/app-quer/

The following is how some issues were approached in the implementation phase:

- The researchers defined a well-defined assigned routine and specified responsibilities during the Quer mobile app installation. And the Joinvasc team was responsible for choosing who would perform the installation of the Quer mobile app on the patient's cell phone.
- Instructions were drafted to help health professionals to explain the installation routine for running the app. That begins by first identifying hospitalized stroke patients, registering the patient on the Web Quer Portal, installing the bedside app in the patient's or companion's cell phone shown in Figure 1 (B), and registering the medications the patient is taking when released from the hospital.
- A cell phone with an activated phone chip number and Wi-Fi access to an internet router has been provided for installing the app for those patients who did not have Wi-Fi access available. That cell phone was also provided to the patient to contact them whenever they had any doubts regarding the research and to operate the app.
- The patients and/or their companions were approached after the patient was transferred to the Stroke Unit, which usually occurs on the third day after hospital admission. Figure 1(A) shows that some information pamphlets were drafted to help the patients and the health professionals, and videos were prepared to explain the Quer mobile app installation and usage.
- A formal introduction to the Quer app took place for most healthcare professionals who provide stroke care at the hospital. It is important to stress that teaching employees to use the Quer app and the Quer Web Portal took place several times during the process; for example, initially, it was explained to the resident nursing students when installing the app, and after that, two nurses who were Joinvasc employees, and then it was explained to all the employees in the U-AVC Integral for inputting the medications in the app.

(A) Information pamphlet for installing the Quer N0 Stroke mobile app.

(B) A Joinvasc nurse teaching the patient's companion how to install the Quer mobile app.

Figure 1. The implementation phase of the 'Quer N0 Stroke' project.

3. Results

In the adaptation phase, three monitoring questionnaires and six stroke healthcare guidance were drafted. Figure 2 displays the main function menus from the Quer app

after downloading the Quer app from the app store, and signing up for an account, and joining to Brazilian Stroke Association.

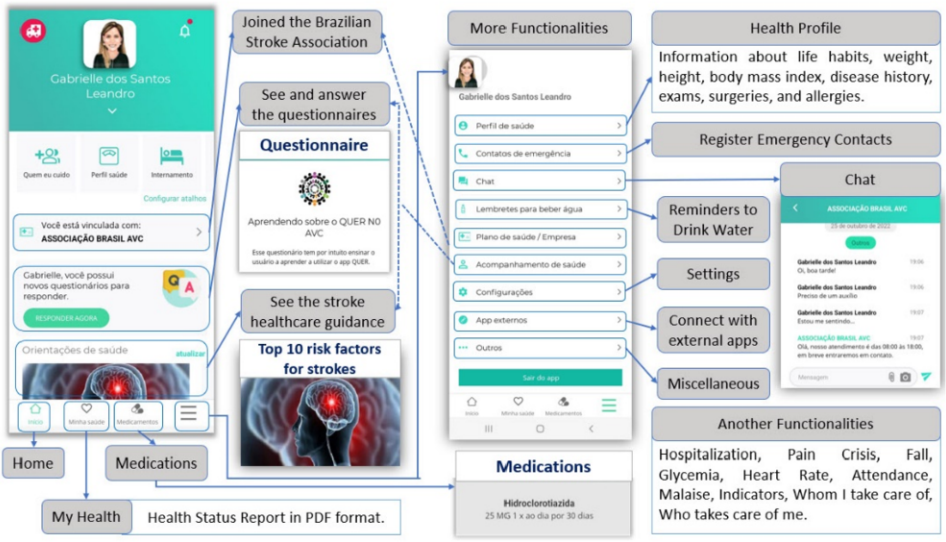


Figure 2. Menus from the Quer mobile app.

In the installation phase, the hospital employee installs the app on the patient's and/or caregiver's cell phone, and the person is taught how to use the Quer app. The information pamphlet on using the app is also delivered, as shown in Figure 1. The healthcare professional will register the patient's medications in the app the patient needs to take before being released from the hospital. That is for helping them to manage the disease (since the app is enabled to register reminders at proper times for taking medications). The patients usually use the chat to clarify their doubts about issues such as rehabilitation, healthy foods, physical activities, and others. The patient can also input data in the app at any time and at the most appropriate time for doing that.

One of the questionnaires named 'Introducing to use Quer mobile app' collected data from 42 patients and identified that 72% of patients had a stroke history in their family (23% father, 28% mother, 26% other family members, and 5% had not one, but two related family members), 86% of the patients were hypertensive, 28% diabetic, 28% had dyslipidemia, and 18% of patient suffered from heart disease. When the issue approached was about personal medical care during the previous year before the stroke, 29% did not take any medical examinations, 36% had one or two medical appointments, 11% had three appointments, and 24% had four or more medical appointments.

The main identified challenges for implementing this technology were the patients needing to have cell phones available for internet access, hindrances in understanding and employing technology, loss and/or stolen cell phones, difficulty in answering the questionnaires, and non-compliance to treatment. However, these issues haven't been approached yet by the researchers.

4. Conclusion

This research portrays adaptation feasibility and the implementation of a cell phone app for monitoring and following up on stroke patients after hospital discharge. The 'Quer' app was proven in this study to be capable of providing some benefits, such as aiding in the adequate use of medication at the prescribed schedules for the patients, facilitating communication with the health professional through the chat, providing stroke healthcare guidance, and automate data collection through questionnaires. Although this app can partially substitute the healthcare data collection process after a stroke, there are still some challenges due to the unavailability of internet access to cell phones and non-adherence to using the app.

References

[1] Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. International Journal of Stroke. 2022;17:18–29.
[2] Akinpelu AO, Gbiri CA. Quality of life of stroke survivors and apparently healthy individuals in southwestern Nigeria. Physiother Theory Pract. 2009;25:14–20.
[3] Gresham GE, Stason WB, Duncan PW. Post-Stroke Rehabilitation [Internet]. DIANE Publishing; 2004 [cited 2022 Feb 19].
[4] Leandro G dos S, Moro C, Miura DY, et al. Process mining leveraging the analysis of patient journey and outcomes: Stroke assistance during the Covid-19 pandemic. Knowledge Management & E-Learning: An International Journal. 2021;421–437.
[5] Leandro G dos S, Miura DY, Safanelli J, et al. Analysis of Stroke Assistance in Covid-19 Pandemic by Process Mining Techniques. 2022.
[6] Prefeitura Municipal de Joinville. Lei no 7.448, de 12 de junho de 2013. Institui o Banco de Dados Registro de Acidente Vascular Cerebral - AVC de Joinville e dá outras providências [Internet]. Diário Oficial Eletrônico do Município de Joinville. Joinville; 2013.
[7] Hughes CML, Padilla A, Hintze A, et al. Developing an mHealth app for post-stroke upper limb rehabilitation: Feedback from US and Ethiopian rehabilitation clinicians. Health Informatics J. 2020;26:1104–1117.
[8] Dhruva SS, Ross JS, Akar JG, et al. Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform. NPJ Digit Med. 2020;3.
[9] McConnell M v., Shcherbina A, Pavlovic A, et al. Feasibility of Obtaining Measures of Lifestyle From a Smartphone App. JAMA Cardiol. 2017;2:67.
[10] Gresham G, Hendifar AE, Spiegel B, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. NPJ Digit Med. 2018;1:27.
[11] GIL AC. Como elaborar projetos de pesquisa. 5th ed. São Paulo: Atlas; 2010.
[12] Prodanov CC, Freitas EC de. Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico [Internet]. 2a. Novo Hamburgo: Feevale; 2013. Available from: www.feevale.br/editora.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230098

Enabling Clinical Trials of Artificial Intelligence: Infrastructure for Heart Failure Predictions

Arian RANJBAR[a,1], Jesper RAVN[a], Erland RONNINGEN[a] and Ole HANSETH[b]
[a] *Medical Technology and E-health, Akershus University Hospital, Norway*
[b] *Department of Informatics, University of Oslo, Norway*

Abstract. The last decade has seen a large increase in artificial intelligence research within healthcare. However, relatively few attempts of clinical trials have been made for such configurations. One of the main challenges arise in the extensive infrastructure necessary, both for development, but particularly to run prospective studies. In this paper, infrastructural requirements are first presented, together with constraints due to underlying production systems. Then, an architectural solution is presented, with the aim of both enabling clinical trials and streamline model development. Specifically, the suggested design is intended for research of heart failure prediction from ECG, but is generalizable to projects using similar data protocols and installed base.

Keywords. Artificial Intelligence, Clinical trials, Infrastructure, Decision support

1. Introduction

The last decade has seen a large increase in research of artificial intelligence (AI)-enabled algorithms within many industries, including health care. Particularly for specialties relying on extensive data during diagnostics, such as radiology and pathology. Initial development, by these modelling techniques, often use performance evaluations limited to retrospective data in laboratory settings, [1]. To realize the potential, and incorporate the models into production, additional clinical testing is necessary; ensuring real world performance, effectiveness and safety. However, there are currently relatively few attempts of such studies and most rely on small populations or non-randomized tests, [2].

Recently the development of electrocardiogram (ECG) classification using machine learning reached desirable performance for various tasks on retrospective data, see e.g. [3]. Similarly, such ECG classification models need further clinical testing before deployment, [4]. Nevertheless, clinical trials require comprehensive preparatory work; both with regards to study design but especially in enabling infrastructure, connecting parts of the underlying health care system. Particularly, allowing prediction models to run in real-time and clinicians to interact with inference results. Then, how can we design an infrastructure that supports efficient development cycles of AI in health care, including clinical trials?

[1] Corresponding Author: Arian Ranjbar, Akershus Universitetssykehus HF, 1478 Lørenskog, Norway; E-mail: arian.ranjbar@ahus.no. This research was funded by Nasjonalforeningen for folkehelsen, and parts of the infrastructure was sponsored by Novartis Norway AS.

This paper aim to first present the infrastructural challenges concerning clinical trials of AI-enabled algorithms, including the precedent stages of model development, within an established healthcare production system. Then, present an infrastructure not only enabling clinical trials but also supporting faster development; by integrating data mining, training and prospective studies into a single platform. Specifically, the presented design is intended for development of heart failure prediction models using ECG, but is generalizable to systems relying on the same design principles.

2. Background

AI-enabled algorithms in healthcare have traditionally been built based on machine learning algorithms, such as neural networks, [5]. Most often, retrospective data is used for both training and evaluation, presuming the recorded data is representative for the total inference population. However, the model may not necessarily be able to extrapolate or generalize outside of the training domain. To prevent deployment of models with limitations in real-world performance, either due to discrepancies in training and inference population, overfitting or other model deficiencies; further prospective studies are carried out for validation, [1].

Prospective studies are generally divided into performance validations of the model or clinical trials, i.e. effectiveness evaluation of the system as e.g. decision support. In the former, after reaching sufficient accuracy on retrospective testing data, the model is queried for inference on prospective data. Such processes can be executed in the background without interfering with the regular production, since involvement of clinicians are unnecessary. Monitoring processes can also be implemented to detect performance drifts, either due to temporal changes or erroneous assumptions of the inference domain. After extensive testing, reaching sufficient performance on both retro- and prospective data, the latter type of study can be implemented. Clinical trials add an additional layer by including a feedback loop between clinicians and the model, to try out the intended interaction. This type of study may also provide metrics for usefulness, efficiency, quality, ease of learning, response time etc., in a real-world setting.

The key enabler to all developing stages is an efficient infrastructure. Model development requires access to relevant databases with the possibility of linking information to build datasets, in the process of data mining. Validation requires data streaming of prospective data and real-time inference capabilities. Finally, the clinical study needs writing access to production systems, without risking interruption of regular activity and patient safety. All requirements need to be delivered while also being compliant with data and privacy regulations, [6].

Designing an architecture that satisfies the requirements, is heavily dependent on the underlying IT infrastructure. The installed base is commonly classified as either a centralized system, via a platform, or distributed system, consisting of IT silos. The former usually hosts a common core with the distinct modules running as applications on top of it, transforming the infrastructural- to a software problem. In this scenario, clinical trials can be executed as additional applications with permission to access relevant data sources. However, most health care systems today run the latter, distributed system; with modules hosted within their own IT silos, often using different standards, [7]. Although not inherently wrong, IT silos generally provide more flexible solutions; the interconnection gets increasingly challenging in this scenario, relying on both hardware and software, [8].

3. Method

An architectural design of a platform for AI development, including clinical trials, is suggested based on the general requirements discussed in the previous section. The platform is specifically intended to be used for heart failure prediction models using ECG data on a distributed installed base. The explicit requirements for the system are:

- Linking retrospective ECG data from a picture archiving and communication system (PACS), [9], with medical records and external data (national medical registers), to do data mining in a safe contained environment.
- Connect external compute infrastructure for model training and inference.
- Enable real time data streaming of incoming ECG data.
- Enable write capabilities into the PACS.
- Prevent the system from causing interruption of regular production and comply with regulation.

4. Infrastructure

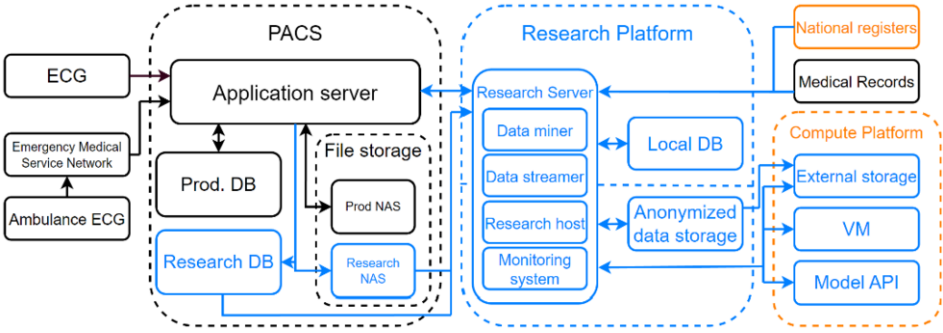


Figure 1. Schematic overview of the suggested infrastructure, arrows indicate data flow. Components of the installed base are represented by black, externally hosted resources are indicated by orange and suggested infrastructure by blue.

The main part of the suggested infrastructure consists of a research platform, serving as an interconnection of the required subsystem hosted within their representative IT silos, see Fig. 1. The PACS handles incoming ECG data, either from the clinics or from ambulances over an emergency medical service network. The network uses the digital imaging and communication in medicine (DICOM)-standard, [10], and data transfer is done via a main PACS application server. Each case is stored within a production database with the respective ECG waveform data in a dedicated file storage. To prevent interruption of the regular production pipeline, a clone of the database and ECG storage is installed. The clones are not only connected to the research platform, but also recipient of external information relayed through the application platform using DICOM-SR. The latter closes the feedback-loop, again, without potentially interfering with production.

The research platform mainly acts as a gateway, running data services such as data mining, anonymization and compute communication. These services can be hosted at two distinct layers with different permissions, where the first layer handles sensitive data and the second layer only has access to anonymized data and non-sensitive information. Therefore, ECG data from the PACS enters the platform at the first layer, together with medical records and externally provided sensitive data. This allows for data mining,

generating training datasets of the recorded retrospective data, necessary for model development. As datasets are generated, indexing is stored in a local database at the first layer before anonymization and exportation to the second layer. At the second layer the platform has access to external compute infrastructure allowing for model training and inference. Finally, as previously mentioned, the research server may communicate directly with the PACS application server, allowing feedback in terms of inference results. This channel also allows the PACS to directly send inference requests to the research platform without going through the PACS database.

5. Implementation

The implementation was done at Akershus University Hospital using ComPACS as PACS, LifeNET as provider for the emergency medical service network and DIPS for medical records. The research platform was hosted using two servers with mirroring for redundancy and real-time patching capabilities. Each service could be hosted in a contained virtual machine with the required permissions. Finally, cloud computing was provided by Google Cloud Platform, connected through the corresponding developer kit, only having access to data storage containing pseudonymized data.

6. Discussion

The presented infrastructure not only enables clinical trials, but streamlines the process from model development to production deployment. Data mining is done directly on the research platform, which is particularly useful if data inclusion needs to be iterated. Prospective studies can run as background processes of the production systems, with real-time monitoring from the research platform, without relying on clinicians relaying relevant cases. Finally, clinical studies are executed in the same environment as development, which is consistent with regular production. This also simplifies the process for clinicians accustomed to the production systems.

6.1. Data regulation

Although the platform has access to the production systems, the data available for processing is strictly limited by data regulation. This is an additional benefit of running clones of the production databases, serving as an intermediate data layer. However, in order to use medical records, both for retrospective and prospective data, relevant permits and ethical assessments needs approval. In particular, due to the consent paradigm, consent either needs to be collected from patients already in the registers, or new data needs to be collected from consenting patients. Such regulations are often adapted to traditional medical research, conforming to data minimization principles, unlike machine learning relying on extensive data, [11].

On this note, Norway recently legislated exemption from the consent paradigm when collection of consent or new data is unfeasible; and the data usage is considered to be of significant benefit and unharmful, in ethical reviews (Helseperonelloven Kap.5 §29). In projects approved for exemption, data mining is significantly easier since the targeted data is directly cloned into the separate storage.

6.2. Limitations

This study investigates infrastructure for clinical trials and efficient AI development cycles, particularly aimed to be used for heart failure prediction modelling. It is also constrained to an installed base using distributed IT silo systems. The suggested platform is thereby tailored for this specific use case. However, these constraints and installed base are commonly used within healthcare, [7]. The solution is also generalizable to projects using similar data setups e.g. data from an archiving system and medical records.

Performing clinical trials of AI-enabled models require a human machine interface (HMI) to provide feedback to clinicians. Detailed examination of HMIs is left out of this study, where PACS is used as communication channel. Although having the benefit of being a regular production system, the effectiveness as mediator is not investigated.

7. Conclusion

This paper investigates efficient designs of infrastructure for AI development in healthcare, enabling clinical trials. Such solutions are heavily dependent on the installed base, but in general needs data mining capabilities from relevant sources, access to compute power, real-time data streaming of prospective data and writing permissions in production systems; all while not risking interference with regular production and complying with data regulation. The suggested solution is intended for development of heart failure prediction models using ECG data, but is generalizable to projects using similar setups. By deploying the platform, the full development cycle is streamlined, from model training to clinical trials. However, further investigations are needed for effective HMI design.

References

[1] Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearbook of medical informatics. 2019;28(01):128-34.
[2] Nagendran M, Chen Y, Lovejoy CA, Gordon AC, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. bmj. 2020;368.
[3] Ribeiro AH, Ribeiro MH, Paixao GM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature communications. 2020;11(1):1-9.
[4] Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. European heart journal. 2021;42(46):4717-30.
[5] Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nature biomedical engineering. 2018;2(10):719-31.
[6] Rumbold JMM, Pierscionek B. The effect of the general data protection regulation on medical research. Journal of medical Internet research. 2017;19(2):e7108.
[7] Bygstad B, Hanseth O. Transforming digital infrastructures through platformization. Association for Information Systems; 2019.
[8] Aanestad M, Grisot M, Hanseth O, Vassilakopoulou P. Information infrastructures and the challenge of the installed base. Information infrastructures within European health care. 2017:25-33.
[9] Huang H, Taira RK. Infrastructure design of a picture archiving and communication system. American Journal of Roentgenology. 1992;158(4):743-9.
[10] Mildenberger P, Eichelberg M, Martin E. Introduction to the DICOM standard. European radiology. 2002;12(4):920-7.
[11] Ranjbar A, Skolt K, Aakenes Vik KT, et al. Fairness in Artificial Intelligence: Regulatory Sandbox Evaluation of Bias Prevention for ECG Classification. Proceedings of Medical Informatics Europe; 2023.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230099

Post Hoc Sample Size Estimation for Deep Learning Architectures for ECG-Classification

Lucas BICKMANN[a,1], Lucas PLAGWITZ[a] and Julian VARGHESE[a]
[a]*Institute of Medical Informatics, University of Münster, Münster, Germany*
ORCiD ID: Lucas Bickmann https://orcid.org/0000-0001-9043-7844, Lucas Plagwitz https://orcid.org/0000-0001-7626-8853, Julian Varghese https://orcid.org/0000-0002-7206-3719

Abstract. Deep Learning architectures for time series require a large number of training samples, however traditional sample size estimation for sufficient model performance is not applicable for machine learning, especially in the field of electrocardiograms (ECGs). This paper outlines a sample size estimation strategy for binary classification problems on ECGs using different deep learning architectures and the large publicly available PTB-XL dataset, which includes 21801 ECG samples. This work evaluates binary classification tasks for Myocardial Infarction (MI), Conduction Disturbance (CD), ST/T Change (STTC), and Sex. All estimations are benchmarked across different architectures, including XResNet, Inception-, XceptionTime and a fully convolutional network (FCN). The results indicate trends for required sample sizes for given tasks and architectures, which can be used as orientation for future ECG studies or feasibility aspects.

Keywords. machine learning, ecg, sample size, estimation, deep learning

1. Introduction

Twelve-lead electrocardiograms (ECGs) are complex time-series which require a large amount of manual expertise and time for annotation. Still, they give insights for many heart malfunctions and diseases. Automating and increasing the classification of these tasks are an important part for the future of ECG-based precision medicine. With the advancement of machine learning in the recent years, many possibilities arise in life-sciences. However, medical datasets are scarce, but machine learning and especially deep learning preferably require large datasets. Additionally, traditional sample size estimations are hardly applicable for machine learning, and it is arduous for the large variety of architectures and different tasks [1][2]. Pre hoc estimates are based on model parameters, but are challenging to compute, especially for the increasing complexity of architectures. This paper aims to introduce a large-scale post hoc sample size estimation on different architectures for binary classification tasks on ECGs, by computing and fitting the learning curve. This gives researchers insights and guidelines for required samples, as well as the ability to estimate a sample size for new studies of automatic ECG-classification.

[1] Corresponding Author: Lucas Bickmann, E-mail: lucas.bickmann@uni-muenster.de.

2. Methods

2.1. Dataset

The training is conducted on the PTB-XL v1.0.2 [3][4][5] with a sampling rate of 100Hz, including 21801 samples. The tasks are binary classification for the PTB-XL diagnostic super-classes Myocardial Infarction (*MI*), Conduction Disturbance (*CD*), ST/T Change (*STTC*), and *sex*. For each classification task, the dataset consists of single labeled *NORM* (healthy controls) and the respective diagnostic superclass. For *sex*-classification, only *NORM* annotated ECGs are selected. The training was conducted on predefined folds 1-8 using shuffled stratified sampling, with a variable number of samples from 100 to a maximum of 4000. To complement the increasing training-samples size, the validation-set increases linearly alongside up to the complete fold at 4000 train-samples. Yet, a minimum of 7.5% of the validation-fold is defined for those train-sample splits which would result in a lower fraction. This estimates dataset splits in real world conditions and constricts the minimal number of samples for a quality validation estimate. The validation- and test-datasets are fold 9 and 10 respectively, as suggested by the authors of PTB-XL, as these include manual curation. The label distribution for the validation and test set are given in Table 1.

Table 1. Normal/anomaly and male/female distribution in the datasets (rounded).

	CD	MI	STTC	Sex
Validation	.187	.255	.278	.972
Test	.202	.281	.265	.880

2.2. Training

The training is conducted with python3.9 using tsai [6]. The chosen architectures are partly based on highest benchmark scores [7]. These include *XResNet1d101*, *InceptionTime*, *XceptionTime* and *FCN*. Each training is conducted 25 times for each number of samples and each architecture, resulting in 225 datapoints for each individual architecture and classification task. The initial learning rate is estimated via an initial run of a learning rate finder [8]. The *1cylce* policy [9] with a maximum number of 500 epochs, early stopping callback with a $\delta = 5e^{-3}$ and a patience of 50 with validation loss as monitoring metric is utilized. The chosen loss function is weighted Cross Entropy. Each sample is standardized independently using batch transformations. The training is conducted on a NVIDIA A40 GPU with a train-batchsize of 1024. Testing and validation are conducted using an equally large batchsize of 1024.

2.3. Evaluation

The half standard deviation around the mean, means (dot-markers) and a logarithmic trendline with $f(x) = a + log_{10}(x) * b$ for each architecture is plotted. A combined average plot is computed via the equally weighted average of all binary classifications. The optimal-threshold-point (x-markers) is given by the highest deviation between the trendline and a linear function, which is computed via the origin and trendline value at 4000 samples, for each architecture respectively. This illustrates the point of maximum score and diminishing returns of the gained performance in relation to extra samples.

3. Results

Figure 1 shows the balanced accuracy score (BACC) for each of the individual selected targets. It stands out that *XceptionTime* is in the two top performers with *InceptionTime*, except for *MI*-classification, for which the latter performs the worst. *XResNet* performs very close, but slightly better to a *FCN* in *CD*, *MI* and *STTC*, while falling behind in *sex*-classification by a small margin. The peak performance is mostly reached for all models at 4000 samples, with an exception. *MI*-classification of *InceptionTime* is stagnating very early at ≈ 1000 samples. The optimal-threshold-point is distributed differently for all architectures and all targets between ≈ 370-1030 train-samples. *XceptionTime* requires most train-samples for all targets, while *InceptionTime* requires the least, except for *CD*-classification.

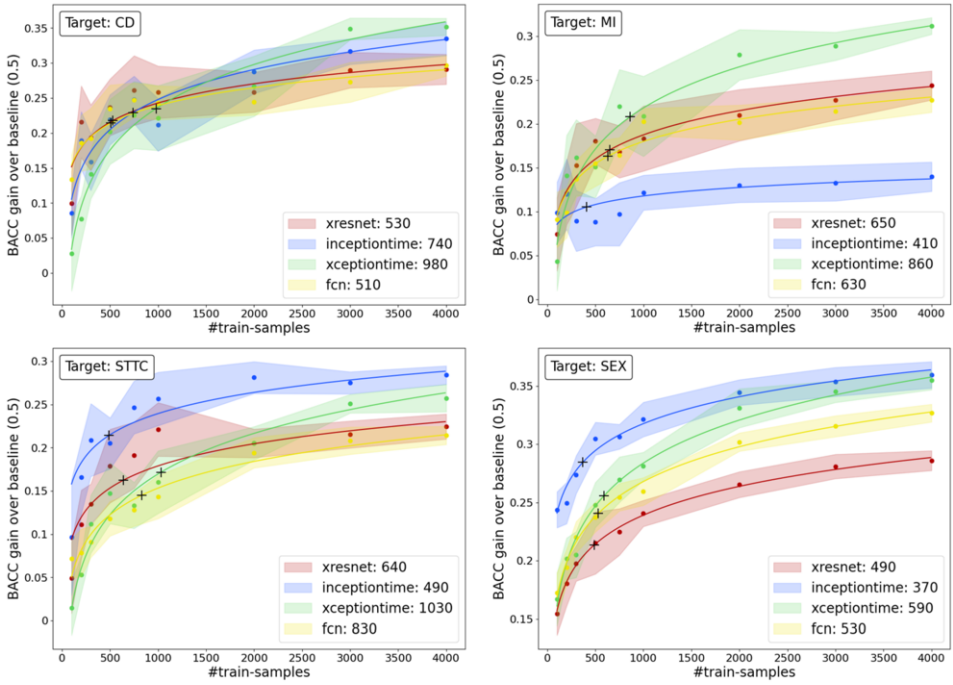


Figure 1. BACC-score gain over baseline of binary classification for different architectures and classification targets: Conduction Disturbance (CD), Myocardial Infarction (MI), ST/T Change (STTC), Sex.

Figure 2 visualizes the combined average performance. It clearly outlines the highest average performing *XceptionTime*. However, the trendline clearly state the necessity of slightly >1000 samples to outperform all other architectures in mean. This fact is underlined with the highest optimal-threshold-point of ≈ 840 train-samples. Yet, *InceptionTime* performs the best in average for ≤ 1000 number of train-samples. At ≈

530-610 samples, all architectures, except *XceptionTime*, perform with the highest BACC/sample-ratio, but still having a rather high standard deviation.

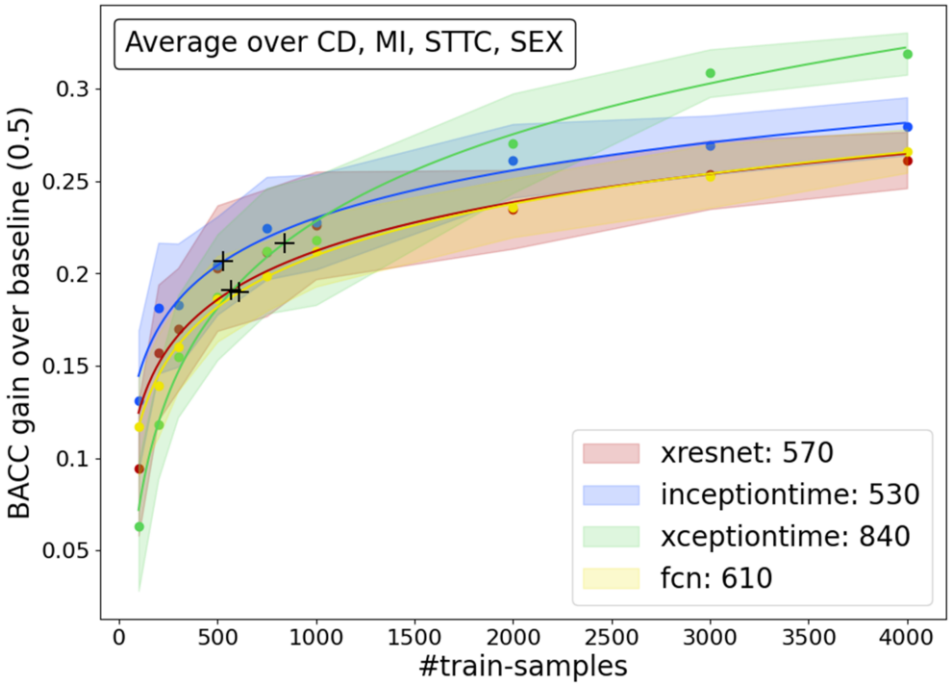


Figure 2. Average BACC-score gain over baseline of binary classification for different architectures.

Table 2 shows the size and label distribution in the validation-set for a train-set with 530 samples. Considering the growing validation-dataset, these additional samples have to be included in the perspective. Therefore, a total of ≥ 650 samples, depending on architecture and target, are required for achieving comparable results, as shown previously.

Table 2. Label distribution in the validation dataset for 530 train-samples (interpolated).

Type	CD	MI	STTC	Sex	Type
Normal	120	120	120	61	Male
Anomaly	22	30	33	59	Female
Total	142	150	153	120	Total

4. Discussion

The highest BACC/sample-ratio is naturally dependent on the classification target, but also on the chosen architecture. Whereas *XceptionTime* excels with a larger number of samples and consistently outperforms other models, *InceptionTime* performs the best in the averaged mean over all targets in the range of 100-1000 train-samples. Yet, it performed worst in *MI*-classification. Therefore, we advise to conduct training with at least two architectures for a specific target, to double-check inconsistencies for the specific task.

5. Conclusion

This paper shows results and a guideline for a preferred minimum number of samples, which yield the highest per-samples-scores. For a lower number of samples (< 1000), *InceptionTime* performs the best, otherwise *XceptionTime* excels. We suggest a minimal train-size of ≥ 530 samples for most applications, to exploit the highest BACC/sample-ratios. The train- and validation-set should therefore contain at least ≈ 650 samples combined. Test-samples are not included in this approximation. We only evaluated binary classification as a first step, and it remains interesting which outline can be drawn for multi-class, -label and regression tasks in future research. Additional apprehension could be drawn for other targets and additional datasets as well. We suggest another study with additional tasks, such as regression, using the same environment and parameters. To summarize, this sample size estimation for binary classification tasks on electrocardiograms indicate helpful guidelines for further research.

References

[1] Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. Canadian Association of Radiologists Journal. 2019;70(4):344-53. Available from: https://doi.org/10.1016/j.carj.2019.06.002.

[2] Du SS, Wang Y, Zhai X, Balakrishnan S, Salakhutdinov RR, Singh A. How Many Samples are Needed to Estimate a Convolutional Neural Network? In: Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc.; 2018. Available from: https://proceedings.neurips.cc/paper/2018/file/03c6b06952c750899bb03d998e631860-Paper.pdf.

[3] Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.2). PhysioNet. 2022 Aug. Available from: https://doi.org/10.13026/zx4k-te85.

[4] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze F, Samek W, et al. PTB-XL, a large publicly available electrocardiography dataset. Sci Data [Internet]. 2020 [cited 2022 Nov 30];7((1):154). Available from: https://www.nature.com/articles/s41597-020-0495-6.

[5] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation [Online]. 2000 Jun;101(23):e215–e220.

[6] Oguiza I. tsai - A state-of-the-art deep learning library for time series and sequential data. 2022. Available from: https://github.com/timeseriesAI/tsai.

[7] Strodthoff N, Wagner P, Schaeffter T, Samek W. "Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL,". IEEE Journal of Biomedical and Health Informatics. 2021 May;25(5):1519-28. Available from: https://doi.org/10.1109/JBHI.2020.3022989.

[8] Smith LN. Cyclical Learning Rates for Training Neural Networks. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV); 2017. p. 464-72.

[9] Smith LN, Topin N. Super-Convergence: Very Fast Training of Residual Networks Using Large Learning Rates. CoRR. 2018 May;abs/1708.07120. Available from: http://arxiv.org/abs/1708.07120.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230100

How to Represent the Patient Voice in the Electronic Health Record?

Line SILSAND[a,1], Gro Hilde SEVERINSEN[a], Gro R BERNTSEN[a] and Aslak STEINSBEKK[a,b]
[a] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
[b] *Department of Public Health and Nursing, Norwegian University of Science and Technology, Norway*

Abstract. There is an agreement among patients, professionals, as well as leaders, and governance that person-centered care (PCC) is central to care quality. PCC care is a sharing of power to ensure that the answer to: "What matters to you?" drives care decisions. Thus, the patient voice needs to be represented in the EHR to support both patients and professionals in the shared decision-making process and enable PCC. The aim of this paper is therefore to investigate how to represent the patient voice in an EHR. This was a qualitative study of a co-design process with six patient-partners and a team of healthcare personnel. The result of the process was a template for the information needed to represent the patients' voice in the EHR based on three questions: "What is important for you right now?". "What matters to you in your life?". "What do you want your care team to know about your history?".

Keywords. Person-centered care, what matters to patients, digital work-tool

1. Introduction

Patients, professionals, leaders of care delivery organizations, and governance alike consider person-centered care (PCC) as central to care quality [1]. Yet there is a significant gap between the rhetoric of PCC and the frontline delivery of PCC. The call for PCC is more urgent now than ever, as system demands on professional conduct leave little room for the enactment of PCC. Professional burnout is rising [2], and patient concerns are lost amid conflicting and rigorous requirements regarding documentation, cost-control, and adherence to guidelines, procedures, and standards. Ironically, the triad of PCC, integrated and proactive care (PIP-care) is thought to be synergistic, and key to the very outcomes systems are seeking [1]. Thus, improving PCC remains a high-level goal, both because it is the right thing to do, and because it is linked to improved patient, professional and system outcomes [3].

"*PCC care is a sharing of power to ensure that the answer to: "What matters to you?" drives care decisions. Patients and professionals work together, within the constraints set by the care system, in a care process to achieve goals that are meaningful to the person.*" [4]. While most interventions on PCC have focused on educating patients and professionals, little attention has been given to the information support directed at

[1] Corresponding Author: Line Silsand, E-mail: line.silsand@ehealthresearch.no

professionals. In the typical outpatient visit, the professional makes decisions on diagnosis and treatment based on the case history, the EHR information, and the clinical exam in the span of 10-15 min. The majority of patients > 60 years are multimorbid, in which case it is even more difficult to get a good understanding of the patient's challenges within the time limits. In addition, there are documentation and communication tasks to serve administrative needs.

In such a time-constrained context, spending time on an interview on the challenging topic of "what matters" to the patient, seems almost unreasonable. Yet, omitting this information may lead to interventions that are misunderstood, poorly adapted to patient circumstances, in contradiction to patient beliefs, or a burden of treatment that is unsurmountable [5, 6]. Especially in cases of complex multimorbidity, rigorous application of guideline recommendations may be disruptive and overwhelming [7]. The sum of these issues leads to loss of patient adherence, and a waste of time and resources spent [5, 6].

Thus, it is important that the patient's voice is represented in the electronic health record (EHR). This can increase clinician awareness and improves the clinicians' opportunity to tailor care to the unique patient. There is, however, a gap of knowledge on how to design a user interface and presentation of the patient's voice in the EHR, which can support both patients and professionals in the shared decision-making process and thereby lower the barrier to patient-centered care.

Our aim was therefore to investigate patient and professional views and suggest how to represent the patient voice in an EHR.

2. Methods

This paper is based on data from a qualitative co-design process with six patients with complex and long-term health problems and next of kin, hereafter named patient-partners, and a team of healthcare personnel working according to the PIP principles [1], in the Patient-Centered Team (PACT) [8]. Data from the patient-partners was collected through semi-structured interviews lasting up to 2 hours and two full-day workshops. The topic for the data collection was how the patient's voice, in terms of needs, values and preferences could be presented in an EHR to support collaboration and person-centered care for patients with complex long-term care needs. We also collected data through a series of four workshops with the PACT professionals. We focused on how to elicit necessary information and document it most efficiently. The workshops were audio-recorded, and summaries were written afterward. In addition, we did two full-day observations of PACT in clinical practice. The data was analyzed thematically in an iterative process aiming to provide a template for the information needed to represent the patient's voice in the EHR [9]. The template content was also discussed within the research group, and with the patient-partners and PACT in several iterations before we arrived at a final template.

3. Results

3.1. The Patient-Partners

The patient-partner wanted to make their health and life stories available for healthcare personnel involved in their services. They envisioned a digital summary of the key events of their health and life journey in form of a text or a video. The summary should enable professionals to see the patient as a whole person "*not just a bundle of diseases*", and as a starting point for meetings between the patients and healthcare personnel.

The patient-partners underlined that their social life, relationship, networks, and role in society shape the way they live with one or more diseases. Therefore, the EHR must focus on more than diseases, tests, and diagnoses. They wish to be understood as individuals in terms of how they cope with their conditions in their everyday life. *"It is important that doctors and nurses know what you have in your 'backpack"* (patient partner).

Another reason for wanting a summary was to avoid having to repeat their story again and again. For multi-morbid patients who have a long complex medical history and/or have lived a life with various other types of challenges and traumas, it is stressful to continuously repeat their stories, especially traumatic experiences. A summary might avoid the reopening of old wounds and reduce the stress of remembering salient details for each new provider.

A third reason was that the summary could be a more effective way of providing new professionals with the information they needed. In the relatively short consultations, the patient-partners experienced there was not time enough to both tell the story and present the current problem and medical examination. Hence, they were forced to focus on the current problem, which led to the omission of relevant information from their past.

Regarding the content of the summary, the patient-partners emphasized an overview of their history, but also included "What matters to you?". By the latter, they meant the summary should outline what they valued in their life and wanted to focus on forward, both in the short and long-term, so that this could be considered in the healthcare plans. To make the summary, the patient-partners suggest some form of an asynchronous dialogue between the patient and healthcare providers to provide the care team with important health-related issues to focus on upfront of an appointment.

3.2. What is Important for Health Professionals to Work with Coordination?

The PACT team, in line with the patient-partners, also talked, about the benefits of having a brief summary of the patient's status and history, the patient voice, for their work. They said that such a summary could be used as the basis of their methodology to ensure coordinated patient-centered care. They especially outlined the importance of identifying the patients' goal by asking "What matters to you? to both understand what was important to the patient and to get to know their personal life history." They translated this goal together with the patient, the next of kin, and other professionals into a plan of actions to help the patient to reach the goals.

To make the summary, the PACT team gathered information through dialogues with the patient and next of kin in addition to a thorough reading of the patient's medical record from different EHR systems to get a full overview of the patient's life. A member from PACT also stated *"[We] have time to do the thorough reading because it is part of our methodology"*. In their experience, other healthcare personnel working in traditional

hospital wards, outpatient clinics, homecare services, or the GPs did not have the time and resources to make such extensive summaries.

They also emphasized that outlining the patients' long-term goals was not a trivial task. In their experience, patients often focused on short-term goals such as getting well enough to go home and manage life on their own. Thus, they had to ask a set of different questions to get patients to talk about long-term life goals. They also experienced that they needed to gain the patient's trust to get them to talk about their lives and what was important for the future. PACT underscored the importance of outlining the actual voice of the patient, in their own words, not just healthcare personnel's observations and interpretations of the patient's situation and what would be best for them.

When asked about how the current EHRs represent the patient's voice, they said that there was a prompt addressing "What matters to you?" in the daily nursing document template at the hospital. However, this field was seldom used, being either deleted or empty, because routines for when to ask and update this question was lacking and how to address the answers they got. "*You cannot ask the patient what matters to you in the same frequent manner as you ask about pain or observing a wound or fluid balance" (PACT member).* They also noted that even if there was a note on "what matters to you", it often disappeared in the long list of documents that evolved continuously, and often it had no link to the actions taken.

3.3. The Template for Digitally Outlining the Patient Voice

The template content included prompts to patients/ professionals that outlined the type of information needed, to cover the salient points in the history according to the patient (past, question three), the current situation (present, question one), and the long-term goals (future, question two). The overall concern was to bring forward information that was pertinent to the shared decision-making process where patients and professionals both contribute.

The first question was "What is important for you right now?" The purpose was to ensure that the short-term needs of the patient were included. As shown above, the PACT team had the experience that these instant goals were most important for the patients in an ongoing care situation.

The second question was "What matters to you in your life?" The purpose of this question was to attend to the more long-term health-related but also overall life goals of the patient in addition to the immediate ones.

The third question was "What do you want your care team to know about your history?" The purpose of this question was to get a summary of the patient's former important health and life events that might impact current care decisions. These include both aspects of life events, and the case history which may include care gaps, unresolved challenges, and expectations not yet fulfilled.

4. Concluding Discussion

We found that the patient-partners, health care personnel, and the researchers agreed upon three prompts based on "what matters to you?" that guide the presentation of the patient voice in the EHR: 1) current wishes, 2) the long-term wishes and 3) the key elements of the case history according to the patient. This "tool" can contribute to a PCC, integrated and proactive care approach, based on the patient's own understanding and

goals for his own life. However, it is not enough to know "what matters" to the patient, the information should also lead to action. Goals can be translated into concrete action plans, which are in turn evaluated against the patient's "what matters". As the patient's situation and goals may change, iterative review and update of "what matters" may also be necessary.

With a stronger presentation of the patient's voice, we show how the patient's answer to "what matters" can be a guiding principle for concrete treatment and follow-up. Further research should investigate how feedback loops from patients regarding patient goals may support improvements of general quality of care and outcomes.

Conclusion: The 'patient's voice' is a prerequisite for shared decision-making and a person-centered approach but is currently missing in the EHR. Thus, the findings in this study can be used to remedy this situation by providing patients and professionals with a practical "tool" that raises awareness, supports implementation, and evaluation of meaningful goals for patients.

Limitations: The 'patient's voice'-approach is mainly feasible work in chronic disease, and not e.g. in acute conditions. It is also intended as something that can be used during a consultation to overcome issues like patients' literacy. Still, some patients could be asked to complete these questions themselves, thus saving the clinicians some time while at the same time strengthening the patient's voice.

References

[1] WHO global strategy on integrated people-centred health services 2016-2026. Executive Summary. Placing people and communities at the centre of health services. World Health Organization; 2015.

[2] Physician burnout: the need to rehumanise health systems. Lancet (London, England). 2019;394(10209):1591. Available from https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32669-8/fulltext#articleInformation

[3] Coulter A, Entwistle, V. A, Eccles A, Ryan S, Shepperd S, Perera R. Personalised care planning for adults with chronic or long-term health conditions. Cochrane Database of Systematic Reviews. 2015;3:1-129.

[4] Berntsen G, Chetty M, Ako-Egbe L, Yaron S, Phan P, Curran C, et al. Person-Centred Care Systems: From Theory to Practice. A White paper for ISQUA.: ISQUA; 2022. Report No.: ISBN 978-0-9955479-2-6.

[5] Weiner SJ, Schwartz A. Contextual errors in medical decision making: overlooked and understudied. Academic Medicine. 2016;91(5):657-62.

[6] Tran V-T, Barnes C, Montori VM, Falissard B, Ravaud P. Taxonomy of the burden of treatment: a multi-country web-based qualitative study of patients with chronic conditions. BMC Med. 2015;13(1):1-15

[7] May C, Montori VM, Mair FS. We need minimally disruptive medicine. BMJ [Internet]. 2009; 339:b2803 doi:10.1136/bmj.b2803. Available from: http://www.bmj.com/cgi/content/full/339/aug11_2/b2803.

[8] Berntsen G, Dalbakk M, Hurley JS, Bergmo T, Solbakken B, Spansvoll L, et al. Person-centred, integrated and pro-active care for multi-morbid elderly with advanced care needs: a propensity score-matched controlled trial. BMC Health Serv Res. 2019;19:682.

[9] Klein H, Myers, M. A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Quarterly. 1999; 23: 67-94.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230101

User Satisfaction with Recently Deployed Electronic Health Records

Ove LINTVEDT[a,b,1], Luis MARCO-RUIZ [a] and Rune PEDERSEN[a]
[a] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
[b]*Nord University, Bodø, Norway*
ORCiD ID: Ove Lintvedt https://orcid.org/0000-0002-4134-4876

Abstract. The high investments in deploying a new Electronic Health Record (EHR) make it necessary to understand its effect on usability (effectiveness, efficiency, and user satisfaction). This paper describes the evaluation process related to user satisfaction over data gathered from three Northern Norway Health Trust hospitals. A questionnaire gathered responses about user satisfaction regarding the newly adopted EHR. A regression model reduces the number of satisfaction items from 15 to nine, where the result represents user EHR Features Satisfaction. The results show positive satisfaction with the newly introduced EHR, a result of proper EHR transition planning and the previous experience of the vendor with the hospitals involved.

Keywords. Electronic Health Records (EHR), usability, user satisfaction, human factors, Computerized Clinical Decision Support Systems (CCDSS)

1. Introduction

Norway's second wave of digitalization in healthcare during the 2010s involved implementing a new Electronic Health Record (EHR) system based on open standards. The requirements for better interoperability, scalability, and information governance [1] took much work to achieve with this first generation of EHRs.

In this regard, in 2012, the Norwegian authorities procured a new EHR that could quickly escalate its information architecture to new domains using open standards and terminologies [2]. The company DIPS was selected to implement a new EHR based on open standards (openEHR) to fulfil these requirements. The investment in the system has been over 90 million euros, which makes it necessary to understand the impact of the system holistically to assess its value for money. Several dimensions are involved in evaluating health information systems [3]. Some of these dimensions are long-term and economic outcomes [4], while others measure the actual value in daily practice [5].

The new EHR, DIPS Arena, was deployed in all hospitals in Northern Norway in 2021. This paper aims to assess the impact of the new EHR on usability, satisfaction, and user experience. The study has a twofold objective: to understand the effect of the EHR transition and to analyse which dimensions of the survey have the highest impact on user

[1] Corresponding Author: Ove Lintvedt, E-mail: ove.lintvedt@ehealthresearch.no, phone +4796519000.

satisfaction for improving the National Implementation-research Network eHealth (NINe) questionnaires for the incoming years.

2. Methods

Setting: We approached clinical users from the newly implemented EHR at the Northern Norway Regional Health Authority: University Hospital of North Norway (UNN), Nordland Hospital (NLSH), and Finnmark Hospital (FSH). All hospitals had been involved in the transition from their previous system, DIPS Classic, into the new system, DIPS Arena.
Data collection: All hospital employees were invited by email with no exclusion criteria and those who did not work with patients was excluded. Of 603 participants, 221 EHR users completed the survey, including 25.8% physicians, 36.2% nurses, and 38.0% from other professions. The hospital sent two reminders from September to December 2021.
Questionnaire: The survey followed the ISO standard (9241-11) [6] of usability: effectiveness, efficiency, and satisfaction. It was developed based on past research using a previously validated questionnaire [7] and measured user satisfaction in three categories: EHR Function satisfaction (11 items; Q1-Q11), EHR Generic satisfaction (four questions; G1-G4), and EHR Overall satisfaction (one question). Login attempts and EHR malfunctions measured interruptions. A 5-point Likert scale ('Completely disagree', 'Partially disagree', 'Neutral', 'Partially agree, 'Completely agree') and the survey tool LimeSurvey (LimeSurvey GmbH, Hamburg, Germany) were used.
Analysis and statistical methods:
For analysis, frequency was used for discrete variables, chi-square test for comparing professions across hospitals, and One-way Analysis of Variance (ANOVA) to compare satisfaction means. The significance level was set at p=.05, and SPSS 25 (IBM Corp., Armond, NY) was used for analysis. Missing values were addressed with the Missingness Completely At Random (MCAR) assumption [8], confirmed with Chi-Square ($\chi 2$= 972.41, df=952, p=.316), and then imputed using Expectation-Maximization (EM) analysis. Due to profession-dependent questions, overall there were n=616 (24.4%) missing values. When adjusting for questions that were not relevant and related to profession or role (users had answered "not applicable"), the relevant portion of missing values for this study was n=321 (12.7%).

3. Results

3.1. Baseline data and interruptions

Of the EHR users who completed the questionnaire, n=221 (82.5%) participated, with 70.1% females, mean experience of 17.4 years (sd=11.0), and mean age of 45.9 years (sd=11.6). Mental health and substance abuse (30.8%), medical (29.4%), surgical (19.0%), and other (20.8%) fields were represented. Physicians, nurses, and other professionals accounted for 25.8% (n=57), 36.2% (n=80), and 38.0% (n=84), respectively. Participants were from FSH (28.5%), NLSH (40.7%), and UNN (30.8%). Login interruptions occurred at a mean rate of 12.5 per day, with system crashes at 3.7 (one every three weeks).

3.2. EHR satisfaction

EHR Function satisfaction. Overall, 52.2% of respondents were satisfied, 34.0% were neutral, and 13.8% were dissatisfied with EHR Function satisfaction (Q1-Q11). Q5 had the highest satisfaction rate (78.7%), and Q11 had the lowest (13.1%). Satisfaction rates for Q4 and Q6 were 71.0% and 72.9%. Q2, Q3, and Q9 have reasonable satisfaction rates, between 55.2% to 62.9%. Q8 and Q11 (54.3% and 38.9%) have the highest dissatisfaction rates, see Figure 1. There were no significant differences in satisfaction rates by profession or hospital.

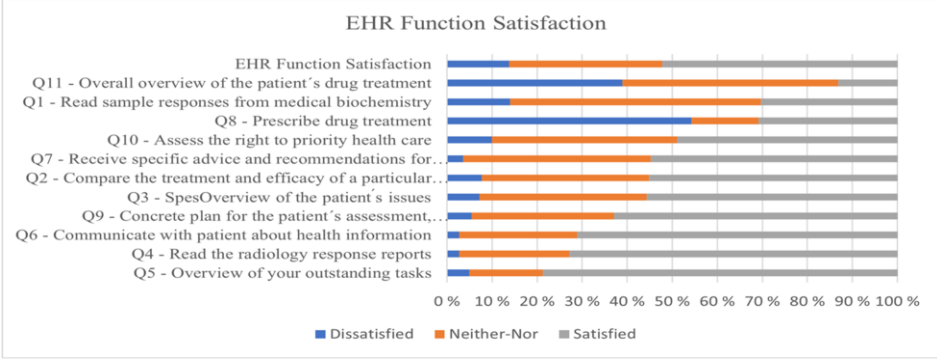


Figure 1. Total and alle items for EHR Function Satisfaction.

EHR Generic satisfaction. Generic satisfaction (G1-G4) measures effectiveness, quality, worth of time and effort, and user-friendliness: 48.1% of respondents were satisfied, 37.7% were neutral, and 14.3% were dissatisfied. Quality received the highest satisfaction rate of over 60%, while user friendliness had the highest dissatisfaction rate of 21.3%, see Figure 2. No significant differences were found by profession or hospital.

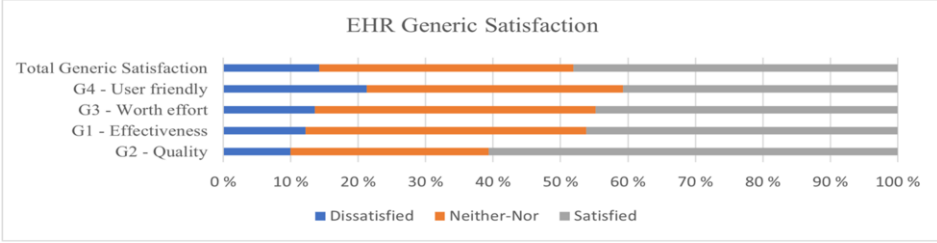


Figure 2. Total and alle items for EHR Generic Satisfaction.

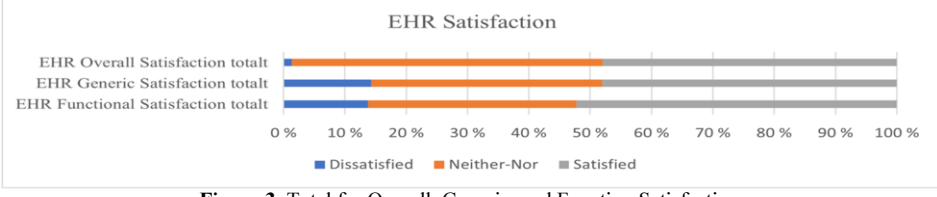


Figure 3. Total for Overall, Generic, and Function Satisfaction.

EHR Overall satisfaction. Overall satisfaction was assessed with a single item, with 48.0% satisfied, 50.7% neutral, and 1.4% dissatisfied, see Figure 3. There were no

significant differences in overall satisfaction by profession or hospital. Satisfaction rates for nurses, physicians, and other professions were 51.2%, 49.1%, and 50.0%, respectively.

3.3. EHR Overall satisfaction predicted by Generic and Function satisfaction.

A regression analysis was conducted to determine whether interruptions, EHR function satisfaction, and EHR generic satisfaction are significant predictors of the EHR overall satisfaction. The normal probability plot from SPSS was inspected and show that normality can be assumed. Both enter and stepwise methods were used to estimate the final model.

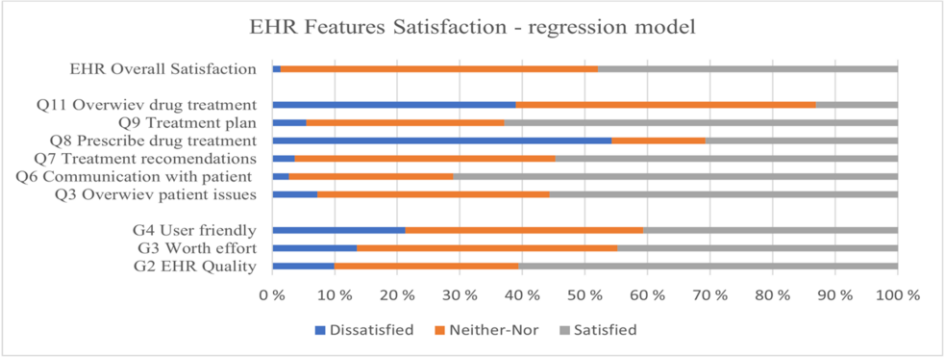


Figure 4. The regression model represents users EHR Features Satisfaction

A linear regression model was estimated correlating all satisfaction items with the overall satisfaction. The model R^2=.83 indicates that 83% of the variance in EHR Overall satisfaction is explained by the reminding nine items: six function satisfaction items (Q3, Q6-Q9, Q11) and three generic satisfaction items (G2-G4), see Figure 4. Interruptions had not significant effect over the EHR overall satisfaction.

4. Discussion

Most users were moderate to highly satisfied with functional, generic, and overall satisfaction of the newly implemented EHR, which is considered a moderate-to-high satisfaction level. Previous studies have reported high dissatisfaction among EHR users in adopting new EHRs [9]. One possible explanation for the moderate-to-high satisfaction level could be attributed to the well-planned implementation and effective handling of incidents during the transition phase. Additionally, the vendor's extensive experience working with these hospitals could have positively contributed to the successful adoption of DIPS Arena into the clinical workflow.

No significant difference was observed among user roles or hospitals. Interruptions did not contribute to the estimated model. Interruptions overlap with other variables that account for the same variance or this could be a response to the new system if the system seems more stable and reduces the number of logins. Some non-significant functional satisfaction items could be due to their specificity. Significant variables in the model are related to more generic functionality (i.e., used by all clinical roles). The Effectiveness

item (G1) was insignificant; further evaluation will be needed. Other items could cover the effectiveness item. Regarding the newly implemented EHR, effectiveness will evolve as users earn experience. The Work Effort item (G3) covers efficiency as the resources used concerning the results achieved.

The nine-item EHR Features Satisfaction model covers a high percentage of the variance in EHR overall satisfaction. However, the model could be further extended to address essential and overarching areas of an EHR satisfaction model following the ISO standard for Ergonomics of human-system interaction.

5. Conclusion

Satisfaction is an essential factor of systems usability measure. The findings in this study, is in line with the ISO 9241-11 usability standard, as we went from three to two satisfaction measures: Features Satisfaction and Overall Satisfaction. The overall high satisfaction rates could be due to a positive experience with the new system. Often, satisfaction will be reduced as the users get more experience with the system, and later, it will increase again. To verify this will require data from the years to come. This remains as future work.

The study authors recommend continuing the study by conducting further research to evaluate the long-term impact of the newly implemented EHR, expanding the satisfaction model, including additional variables, using a mixed-method approach, regularly evaluating user satisfaction to ensure the success of EHR implementation, and include user responses to make results more beneficial for the vendors.

References

[1] European Commission, Directorate-General for the Information Society and Media, Virtanen M, Ustun B, Rodrigues J, Surjan G, Rector A, Stroetmann V, et al. Semantic interoperability for better health and safer healthcare: deployment and research roadmap for Europe. Publications Office; 2013. Available from: doi/10.2759/38514

[2] Ministry of Health and Care Services. Whitepaper. no. 9 (2012-2013): ´One citizen – one Health Record´. St.Meld. nr. 9 (2012–2013) ´Én innbygger - én journal´. Nov. 27, 2012. Available from: https://www.regjeringen.no/no/dokumenter/meld-st-9- 20122013/id708609/ , accessed Mar. 4, 2023.

[3] Nasjonal IKT, Tiltak 48 (2012). Klinisk Dokumentasjon for Oversikt Og Læring, ed, 2009. Available from: http://www.nasjonalikt.no/no/dokumenter/andre_publikasjoner/ , accessed Mar. 4, 2023.

[4] Bassi J, Lau F. Measuring value for money: a scoping review on economic evaluation of health information systems, J Am Med Inform Assoc. 2013 Jul; 20(4): 792–801. doi: 10.1136/amiajnl-2012-001422

[5] West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions, J. Intern. Med., vol. 283, no. 6, pp. 516–529, Jun. 2018, doi: 10.1111/joim.12752.

[6] ISO. Ergonomics of human-system interaction: part 11: usability: definitions and concepts (ISO/DIS 9241-11.2:2016). 2016.

[7] Lintvedt OK, Tayefi M, Nordheim E, Pedersen R, Nicolaisen KM, Lærum H. Electronic Health Records User Experiences: a Nationwide Survey From Norwegian Hospitals. In: Procidings of the eTELEMED 2022, The Fourteenth International Conference on eHealth, Telemedicine, and Social Medicine; 2022, Porto: IARIA. ISBN: 978-1-61208-984-3

[8] Little RJA. A test of Missing Completely at Random for multivariate data with missing values, Journal of the American Statistical Association, vol. 83, pp. 1198-120, Des. 1988, doi: 10.2307/2290157

[9] Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. BMC Med Inform Decis Mak. Feb 2012, 24;12:10.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230102

Conception and Development of a Targeted Alert System : Multisystem Considerations

Frederic EHRLER[a,1] and Katherine BLONDON[a,b]
[a] *University Hospitals of Geneva, Switzerland*
[b] *University of Geneva, Swtizerland*

Abstract. Alerting systems have a strong potential to improve quality of care in hospital by ensuring that clinicians provide more effective and timely care to their patients. Many systems have been implemented but often fail to unleash their full potential due to the problem of alert fatigue. As an attempt to reduce this fatigue we have developed a targeted alerting system ensuring only the concerned clinicians receives the alerts. The conception of the system went through several steps going from the identification of the requirement, the prototyping and implementation into several systems. The results present the different parameters taken into consideration and developed frontends. We finally discuss the important considerations of alerting system, such as the necessity of a governance. The system still needs a formal evaluation to validate that it responds to its promises before being deployed more largely.

Keywords. Alert, Alarm fatigue; Alarm safety, Clinical decision support system, mobile

1. Introduction

Medical alert is a clinical decision support systems that has raise increasing interest in recent years. By helping clinicians provide more targeted and timely care to their patients it improves quality of care. Alert systems can be used to automatically detect and alert clinicians about abnormal laboratory test results, for example, allowing the clinician to react before the patient's condition deteriorates. Medical alerts have the potential to improve patient outcomes by providing clinicians with timely and relevant notifications about their patients' health [1]. For instance, an alert system may also be used to detect a decrease in a patient's blood pressure, alerting the clinician so that they can take immediate action to prevent any further decline. Similarly, an alert system may be used to detect a high level of creatinine in a patient's blood, alerting the clinician to change management plans (e.g., change medications if needed or increase hydration) to avoid further kidney damage [2].

Despite the potential benefits of medical alert systems, there are several barriers that limit their deployment in healthcare. The most significant barrier is that clinicians are often overwhelmed with the sheer number of alerts they receive, which can lead to alert fatigue. This is caused by the fact that traditional alert systems are often configured to produce alerts without targeting a specific user, resulting in many alerts for everyone.

[1] Corresponding Author, Frederic Ehrler, University Hospitals of Geneva, Geneva, Switzerland; E-mail: Frederic.ehrler@hcuge.ch.

Many of these alerts may be irrelevant to most receivers, and lead to a lower responsiveness to alerts overall. In addition, many alert systems are binary (e.g., a result has arrived in the system) but may not have nuances to allow for distinctions between very abnormal findings, somewhat abnormal findings and normal findings. Lastly, many alert systems lack the ability to customize the content of the alert message, making it difficult for the clinician to quickly identify the most urgent matter [3]. In order to effectively address these barriers, a new medical alert system trigger has been developed in our hospital, which in its pilot version is specifically designed to detect and alert clinicians of abnormal laboratory test results. This system is designed to be highly customizable and to consider the specificity of each care unit. It is also designed to reduce alert fatigue by only sending alerts to the clinician in charge when the results of the laboratory tests are above or under a given threshold. Furthermore, the alert message is designed to be concise and easy to read, so that the clinician can quickly identify the most important information.

The purpose of this article is to present the design of a new medical alert system, which is triggered by abnormal laboratory test results. The system is designed to be highly customizable and tailored to target specific members of the patient's medical team.

2. Methodology

To design the alerting system, we went through the following steps: Identification of the goal of the medical alert system, identification of the stakeholders, conduct of a business analysis of the medical alert system. This analysis included the identification of the user needs, the system objectives, the operational requirements, the required features, the system design and the implementation plan. We developed a system design based on business analysis including consideration of the system architecture, the user interface, the data architecture, the security measures and the system integration. We developed a prototype of the system to test the system design and to identify any potential problems or issues. Then we developed a testing plan to test the system and to ensure that it fulfills the requirements of the medical alert system. Finally, we implemented the system once the system design and the testing plan were developed and approved.

3. Results

The goal of the medical alert system is to provide timely and accurate alerts to the clinician in charge of a patient. The stakeholders of the medical alert system include the healthcare professionals (doctor and nurse team), the IT professionals and the software developers.

During the business analysis, we began by identifying all the parameters and attributes that needed to be considered for each alert. These are presented in Table 1. Attributes of the alerts included the level of urgency of the alert (low, medium, high), the need for a recall and the targeted receiver(s). The level of urgency of the alert defines the communication method for the alert: low urgency alerts can wait for the user to open the patient's chart, for example, whereas high urgency alerts need to reach the receiver even when they are not using a computer, such as a smartphone notification. Parameters to define were the thresholds for each laboratory result, for each level of urgency. Other considerations include the text in the alert, its duration of validity (which may be longer

in ambulatory care, and shorter during a hospital stay for example). For alerts with recalls, we also needed to define whether it was a duration of alert or a specific action that would turn the alert and its recalls off.

Table 1. parameter considered for the parametrization of the alerts

Alert parameter	Description	Possible values
Name	Name of the alert	Free text
Description	Description of the alert	Free text
Category	Category of the alert	Clinical / admin / pathways
Urgency	Urgency of the alert	Low / medium / high
Color	Color of the alert	Grey, Yellow, Red
Trigger	What triggers the alert	Threshold
Termination	What switches the alert from active to inactive?	Prescription of anti-Xa HNF activity for heparin alert
Mobile	Is the alert displayed on mobile device	True/false
Displayed information	Information to interpret the result	Free text
Display duration	Duration of display of the alert	Number
Authorized people	Authorized people to handle the alert	Identities of the active directory
Targeted people	Unit receiving the alert	Clincian /nurses / … Unit

The desired process was mapped using business process modeling notation (BPMN) to ensure that all the possibilities were considered. At the start of the shift, doctors identify which patients they will manage (in-patients); for out-patients, the consultation calendar defines the doctor in charge. For this paper, we will focus on in-patients with lab results.

The arrival of a laboratory result triggers the alert: it appears in the medical chart for low urgencies, whereas for urgent, critical results, the doctor is also notified directly on the Bedside app. In specific cases of high risk, the alert is designed with a recall system, if a certain action is not performed. In the case of highly overdosed heparin, it is the prescription of the next anti-Xa activity lab that switches the system off: in our system, each change of dose of heparin (except for the stop) is accompanied by the prescription of the monitoring anti-Xa lab. If no lab is prescribed after an hour, the recall system kicks in, and even alerts the supervising physician, to help ensure that the overdosed heparin is addressed.

In order to make this process actionable, the system is composed of four components that must work altogether. First a rule-based engine, developed with Java Springboot, that can be parametrized, responsible to trigger Kafka event will be consumed in several endpoints; a mobile application, developed in angular, enabling doctor to identify their patients and that receive push notifications; an alert panel in the EHR summarizing the alerts, and finally a header highlighting the most important information. All communications between the components are done using Kafka messages as well as through API structured following FHIR data model.

3.1. The EHR header

All notifications for a given patient are visible in the header of the EHR. The color of the number indicates the highest degree of urgency for a given category of alerts, whereas the number represents the total number of notifications for the patient. In this example we see three categories of alerts, the first of which is the current alerts (left), with at least one important (yellow) alert. The second category indicates that the patient

is in a clinical pathway with one information alert (not urgent, grey). Finally, the third category indicates an urgent administrative alert (e.g., no insurance coverage).

Figure 1. Header displayed on the EHR

3.2. The EHR alert panel

A panel has been developed in our EHR to gather all alerts for a given patient in one place. It opens when the header is clicked, and shows all alerts, sorted by degree of urgency (which is also color-coded). In this panel, users can view the alerts, but also indicate when the alert is addressed: in this case, the alert is moved to the history of sections alerts, at the bottom of the panel. This history is designed for all the times users click "too quickly" on an alert to close it and did not take time to read it!

3.3. The Bedside mobility app

The development of a mobile application was a prerequisite for the good functioning of the system and is the communication channel for targeted urgent alerts. It allows the doctors to define the patients they will manage during their shift, and presents the alerts, which can be seen by clicking on the patient. In this view, the user can choose to hide the alert or not. As mentioned above, alerts can all be found in the EHR panel of the patient's chart.

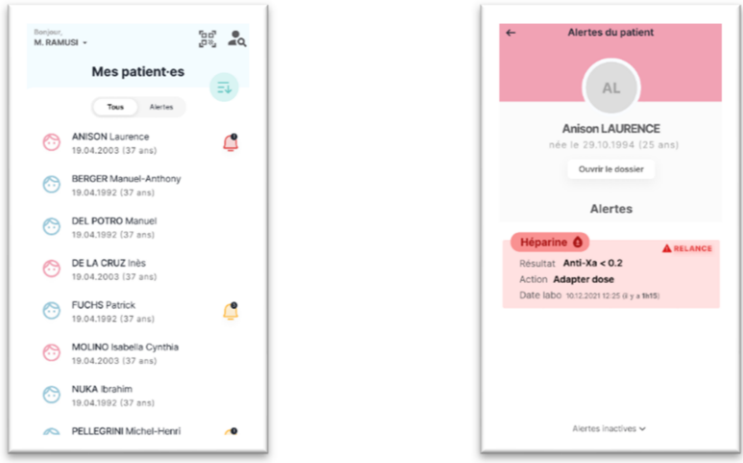


Figure 2. On the left, screen for the doctor to select the patients managed during a shift. On the right, display of an urgent alert (red) about a lab result (blood drop icon)

4. Discussion

Governance for alerts at an institutional level is required to maintain coherence in the system. Closely related alerts (e.g., for different laboratory results) can share many processes, thereby facilitating their implementation. In this case, it is more about getting

the different medical specialties to agree on common thresholds for the institution, or to define division-level thresholds. Governance may need to intervene to reach this consensus [4]. There are often several thresholds to define: besides the upper (and/or lower) threshold of normality, there may also be a threshold for critical values. This exists in the HL7 codes for laboratory results (OBX-8).

However, when individual alerts require specific processes (e.g., recall system for anti-Xa activity for heparin), additional developments are required. In these cases, besides validating the need for these specificities, the governance also defines the priority of the developments. It also has a role in the choice of the solution, which should be the most generalizable for other future alerts, besides addressing safety issues.

Alert fatigue in healthcare is a growing problem which is caused by many alerts being generated in the healthcare system. These alerts are often not targeted, and clinicians are bombarded by a high volume of alerts, leading to alert fatigue. Alert fatigue is a state in which clinicians become desensitized to alerts, leading to them missing important alerts or ignoring them altogether [5]. This can have dangerous consequences, as important alerts could be missed and lead to patient harm. Reducing alert fatigue in healthcare is an important issue that needs to be addressed in order to ensure patient safety. In our project, the identification of the clinician in charge is a key factor to ensure that the right person receives the relevant information at the right time.

5. Conclusion

Alert systems have a strong potential to improve care by ensuring that clinicians provide more targeted and timely care to their patients. One of the strong barriers to these systems are the fatigue induced by alerts. As an attempt to respond to this challenge, we designed a system with several urgency levels and subsequent communication channels, targeting the clinician in charge with messages to help prioritize tasks. Its adoption as well as efficiency must still be evaluated in practice. The pilot system for laboratory results will be generalized to other types of information (e.g., vital signs, administrative data, or clinical itineraries) in the future development of the alert system.

References

[1] Chien SC, Chen YL, Chien CH, Chin YP, Yoon CH, Chen CY, Yang HC, Li YC. Alerts in Clinical Decision Support Systems (CDSS): A Bibliometric Review and Content Analysis. InHealthcare 2022 Mar 23 (Vol. 10, No. 4, p. 601). MDPI. doi:10.3390/healthcare10040601.

[2] Laing C. On the alert for outcome improvement in acute kidney injury. The Lancet. 2015 May 16;385(9981):1924-6.

[3] Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. Journal of the American Medical Informatics Association. 2019 Oct;26(10):1141-9. doi:10.1093/jamia/ocz095.

[4] McGreevey III JD, Mallozzi CP, Perkins RM, Shelov E, Schreiber R. Reducing alert burden in electronic health records: state of the art recommendations from four health systems. Applied clinical informatics. 2020 Jan;11(01):001-12. doi:10.1055/s-0039-3402715.

[5] Joseph AL, Borycki EM, Kushniruk AW. Alert fatigue and errors caused by technology: A scoping review and introduction to the flow of cognitive processing model. Knowledge management & e-learning. 2021;13(4):500.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230103

Why Are Data Missing in Clinical Data Warehouses? A Simulation Study of How Data Are Processed (and Can Be Lost)

Sonia PRIOU[a,1], Guillaume LAME[a], Marija JANKOVIC[a], Gilles CHATELLIER[b], Romain BEY[c], Christophe TOURNIGAND[d], Christel DANIEL[c,e], and Emmanuelle KEMPF[d,e]

[a] *Université Paris-Saclay, CentraleSupélec, Laboratoire Génie Industriel, Fr*
[b] *Université de Paris, APHP-CUP, Department of Medical Informatics, Paris, Fr*
[c] *AP-HP, Innovation and Data, IT Department, Paris, Fr*
[d] *Université Paris Est Créteil, AP–HP, Department of medical Oncology, Henri Mondor and Albert Chenevier Teaching Hospital, Créteil*
[e] *Sorbonne Université, LIMICS, Paris, Fr*

Abstract. In recent years, the development of clinical data warehouses (CDW) has put Electronic Health Records (EHR) data in the spotlight. More and more innovative technologies for healthcare are based on these EHR data. However, quality assessments on EHR data are fundamental to gain confidence in the performances of new technologies. The infrastructure developed to access EHR data - CDW - can affect EHR data quality but its impact is difficult to measure. We conducted a simulation on the Assistance Publique – Hôpitaux de Paris (AP-HP) infrastructure to assess how a study on breast cancer care pathways could be affected by the complexity of the data flows between the AP-HP Hospital Information System, the CDW, and the analysis platform. A model of the data flows was developed. We retraced the flows of specific data elements for a simulated cohort of 1,000 patients. We estimated that 756 [743;770] and 423 [367;483] patients had all the data elements necessary to reconstruct the care pathway in the analysis platform in the "best case" scenarios (losses affect the same patients) and in a random distribution scenario (losses affect patients at random), respectively.

Keywords. Clinical data warehouse, EHR data, data quality, simulation

1. Introduction

Real-world data, and especially Electronic Health Records (EHR) data, are increasingly used to develop innovative digital technologies and new services supporting various activities of health professionals, in research, care and training. To integrate data from multiple sources and provide associated services for secondary use of this data, hospitals have recently started to develop Clinical Data Warehouses (CDW) [1]. However, EHR data are not without flaws, and caveats and recommendations have been raised when re-using them for research [2,3].

[1] Corresponding Author: Sonia PRIOU, Université Paris-Saclay, CentraleSupelec, Laboratoire Génie Industriel, Gif-sur-Yvette, France; Email: sonia.priou@centralesupelec.fr

A major aspect of re-using real-world data is data quality assessment. For EHR data, it doesn't just concern the data entered in the Hospital Information System (HIS). Issues can occur and data can be lost or somehow modified during the extraction, the transformation, or the loading phases, ultimately affecting the dataset provided to researchers [4,5]. This could lead to a loss of power or to systematic bias, by shifting or reversing the results of analyses. Since CDWs combine data from multiple pieces of software, many of them proprietary, they ultimately resemble black boxes, and tracing back how data were processed is hard. One can manually compare the data accessible in the CDW with the data in the HIS [6], but this procedure is time-consuming, and impossible when CDW data are de-identified.

The objective of this paper is to model and simulate the effect of the complex structure of a CDW on the data extracted from the HIS. To illustrate this simulation, we will place ourselves in the position of researchers trying to reconstruct breast cancer care pathways using data available in the Assistance Publique – Hôpitaux de Paris (AP-HP) CDW. By simulating a cohort of breast cancer patients treated at the AP-HP whose data is entered in the HIS, we estimate the proportion of the care pathways that can be fully reconstructed using data available in the analysis platform to researchers via the CDW.

2. Methods

2.1. The Hospital Information System

The AP-HP HIS is composed on a main EHR software and numerous specific software dedicated to specialised medical fields (e.g., imaging, pathology, laboratory tests results…). The main EHR software was not installed in all hospitals at the same time, leading to different EHR software being used across the AP-HP at first. As well, specific software varies from hospital to hospital and are prone to new versions. We conducted an inventory of which software was used each year for each hospital of the AP-HP. We hypothesized that, for a given hospital, the software used was consistent during the entire year and across all departments of the hospital.

The data entered in the HIS is composed of different elements: reports (consultation, hospitalisation, meetings…), laboratory results, procedure and diagnostic codes… Each of these data elements is entered in the HIS via either the main EHR software or one (or more) specific software, depending on both hospital and year.

2.2. Modeling care pathway reconstruction

We considered the following standard care pathway, common to all hospitals included in the study: first consultation of a hospital specialist, cancer diagnosis, therapeutic strategy choice, and administration of treatment. At each step of this care pathway, distinct data elements are recorded in the HIS by clinical staff.

We identified the key information needed to reconstruct the standard care pathway and selected the minimal dataset of data elements to find them (figure 1). To identify the first consultation related to the patient's cancer diagnosis, the date and the purpose of the visit are needed. Both are available in the consultation report. The cancer diagnosis and extension status are made by pathology examination and Positron Emission Tomography (PET) imaging, respectively. The reports give information about the diagnosis and the procedure date is linked to the procedure code. The choice of the therapeutic strategy is

made during the Multi-Disciplinary Meeting (MDM). The content and date of the MDM are available in the MDM report. Finally, the treatment administered is obtained through the following structured data: procedure and diagnostic codes, and laboratory results.

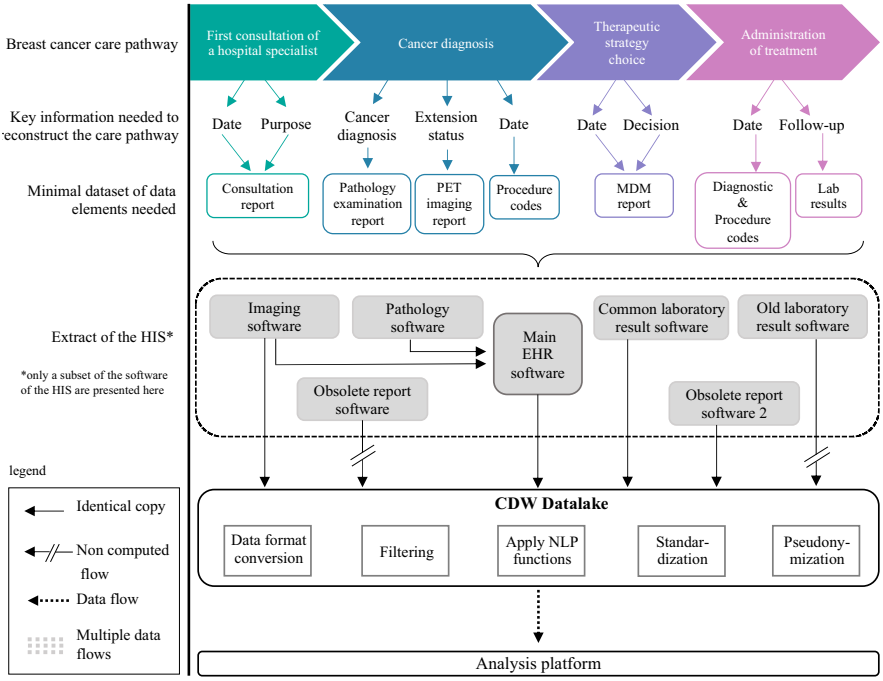


Figure 1. Simplified diagram of the data flow between the HIS and the analysis platform for the standard breast cancer care pathway

2.3. Modeling data processing

Thanks to the reading of internal technical documentation and the help from CDW professionals, we modelled the structure of the CDW. For each data element of the identified minimum dataset, we modelled its trajectory: entry in the HIS, transfer in the CDW and provision on the analysis platform. The trajectory followed by the data elements comprises several steps, between which data is transferred from one element to the next. Each data transfer was qualified according to the type of transformation applied: none, data format conversion, filtering, application of NLP functions, standardization and pseudonymization. Each of these steps was associated with a probability of correct data transfer: in case of impossible transfer, this probability was set to 0; in case of an identical copy, the probability was set to 1; for all other transformations, we modelled the probability of success using a uniform distribution between 0.95 and 1 (figure 1). A sensitivity analysis was performed on the lower bound of the uniform distribution.

2.4. Simulating a cohort of patients

We simulated a cohort of 1,000 patients treated at AP-HP for a breast cancer between 2018 and 2021. At AP-HP, breast cancer is mostly treated in 5 hospitals. One of these

hospitals uses a completely different EHR software than the other 4, so none of its data is integrated into the AP-HP CDW. For that reason, we decided to focus only on the other 4 hospitals, that we will name hospitals A, B, C, and D. We distributed our 1,000 patients across these 4 hospitals and per year following empirical distributions reproducing real proportions in the context of AP-HP. We hypothesised that patients stay in the same hospital during their care and that their entire care pathway occurs during the same year.

We used Monte-Carlo simulation on the success rate of data flows and simulated the success rate coefficients 10,000 times. Each time, we replayed the data trajectory to measure which data elements were available in the analysis platform. We measured the median and inter-quartile range of the percentage of patients for whom all the data elements necessary to rebuild the care pathway were available in two scenarios:

- the "best case" scenario, where all missing elements are concentrated on the same patients (i.e., the same subset of patients are missing their laboratory results, consultation reports, imaging reports…),
- the "random distribution" scenario, where data losses per data elements are randomly assigned to patients (i.e., missing laboratory results are attributed to a random subset of patients; missing consultation reports are attributed to another random subset on patients, possibly intersecting with the previous one).

3. Results

For all hospitals, the median success rate for transmitting the original data from the HIS, through the multiple layers of the CDW and to the analysis platform was 90% [88 – 92] for PET imaging and pathology reports, and 93% [91 – 94] for other reports. The success rate for transmitting procedure and diagnostic codes is 93% [91 – 94] cases. For hospitals using the same laboratory results software, the median success rate for transmitting information to the analysis platform reached 93% [91 – 94]. For hospital A using another laboratory result software, no laboratory results are available in the analysis platform (the data flow does not exist, see figure 1).

Out of 1000 patients, in the "best case" scenario, 756 [743;770] have all the data elements necessary to rebuild their care pathway. In the "random distribution" scenario, 423 [367;483] pathways can be reconstructed (figure 3). Without considering laboratory results, 904 [884;923] and 556 [493;635] pathways can be reconstructed in the "best case" and "random distribution" scenarios, respectively.

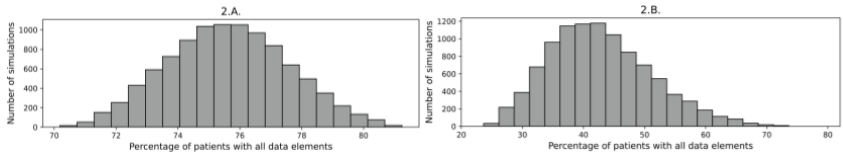


Figure 2. Percentage of pathways that can be fully reconstructed in the "best case" scenario (2.A) and the "random distribution" scenario (2.B)

The sensitivity analysis on the lower bound of the uniform distribution (between 0.90 and 0.99) shows a variation of the number of patients with all data elements available between 700 [673;726] and 804 [801;807], in the "best case" scenario, and 213 [160;282] and 717 [697;736] in the "random distribution" scenario (figure 3).

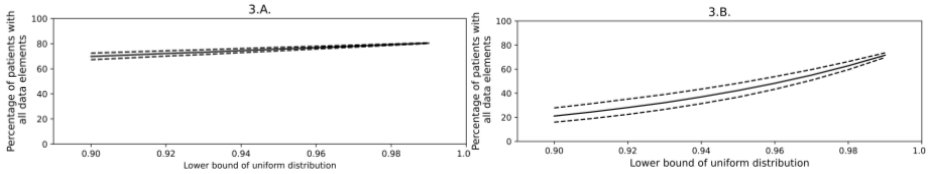


Figure 3. Sensitivity analysis on the lower bound coefficient of the uniform distribution in the "best case" Scenario (3.A) and the "random distribution" scenario (3.B)

4. Discussion and Conclusion

When working with data from a CDW, it is quite common to end up with data completeness issues where we did not expect to find any. This study highlights that, even when each data flow has a high success rate, the percentage of data available at the end can be surprisingly low. Each data element being entered in an independent software, the losses do not necessarily concern the same patients. In this configuration, the more data elements needed for a study, the more patients are going to be excluded because of missing data. More and more studies on data from CDW mention the ratio of patients for which a data element was found. As MDM are required by French law for all patients, one would expect to approach 100%. However, an AP-HP study on colorectal cancer showed that an MDM report was found for only 85% of patients [7].

This study enables to better understand the system behind CDW and the processes through which data are made available. CDW used for research purposes should not be "black boxes", as missing data may be source of biased or imprecise results. In this simulation, every data element was originally entered in the HIS. However, for real life data, the difficulty is not being able to explain the cause of the missing data. For example, is a report missing because it was not written, because it was never entered into the HIS or because of a malfunctioning data flow?

In conclusion, otherwise acceptable performance of individual applications translates to unacceptable overall performance when combined. The next step is to assess each level of data flow risk to better prioritize current and future data flows in the CDW.

References

[1] Tute E, Steiner J. Modeling of ETL-Processes and Processed Information in Clinical Data Warehousing.
[2] Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL, et al. What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask. J Med Internet Res. 2021 Mar 2;23(3):e22219.
[3] Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. Medical Care. 2013 Aug;51(Supplement 8Suppl 3):S30–7.
[4] Ong T, Pradhananga R, Holve E, Kahn MG. A Framework for Classification of Electronic Health Data Extraction-Transformation-Loading Challenges in Data Network Participation. 2017;5(1):16.
[5] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018 May 29;20(5):e185.
[6] Denney MJ, Long DM, Armistead MG, Anderson JL, Conway BN. Validating the extract, transform, load process used to populate a large clinical research database. International Journal of Medical Informatics. 2016 Oct;94:271–4.
[7] Kempf E, Priou S, Lamé G, Daniel C, Bellamine A, Sommacale D, et al. Impact of two waves of Sars-Cov2 outbreak on the number, clinical presentation, care trajectories and survival of patients newly referred for a colorectal cancer: A French multicentric cohort study from a large group of university hospitals. Intl Journal of Cancer. 2022 May 15;150(10):1609–18.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230104

Nationally Shared Medication Lists – Describing Systems in the Nordic Countries

Tora HAMMAR[a,1], Trine S BERGMO[b], Cille BÜLOW[c], Stine S CLAUSEN[d], Unn Sollid MANSKOW[b], Johanna TIMONEN[e] and Anette Vik JØSENDAL[b]
[a] *eHealth Institute, Linnaeus University, Kalmar, Sweden*
[b] *Norwegian Centre for E-health Research, Tromsø, Norway*
[c] *Dept. of Clin. Pharmacol.,Bispebjerg University Hospital, Copenhagen, Denmark*
[d] *Section of General Practice, University of Copenhagen, Denmark*
[e] *School of Pharmacy, University of Eastern Finland, Kuopio, Finland*
ORCiD ID: Tora HAMMAR https://orcid.org/0000-0003-1549-2469

Abstract. This paper provides an overview of shared medication lists (SMLs) in four Nordic countries (Denmark, Finland, Norway and Sweden) with a focus on the type of information the list is based on. This is a structured comparison conducted in stages using an expert group, grey papers, unpublished materials, web pages, as well as scientific papers. Denmark and Finland have implemented their solutions for an SML and Norway and Sweden are working on the implementation of their solution. Denmark and Norway have or are aiming at a list based on medication orders, while Finland and Sweden have lists based on prescriptions.

Keywords. shared medication list, medication, informatics, e-prescription

1. Introduction

Drug-related problems can often be linked to information management issues such as discrepancies between medication lists for the same patient [1-3]. Having an electronically shared medication list (SML) can facilitate safe and efficient medication management and there have been several initiatives to implement such lists [4-6]. Implementing an SML is highly complex, and there is limited knowledge about whether such solutions will have the expected benefits [7]. Four Nordic countries, Denmark, Finland, Norway and Sweden, are all implementing an SML [8-11]. The aim of this paper is to provide an overview of SML solutions in these countries with a focus on what information the list is based on. Increased knowledge about the SML solutions using common terminology can facilitate learning possibilities and future studies comparing effects and identify barriers and facilitators for the implementation of SMLs.

[1] Corresponding Author: Tora Hammar, eHealth Institute, Department of medicine and optometry, Linnaeus University, 391 82 Kalmar, Sweden, E-mail: tora.hammar@lnu.se.

2. Materials and Method

This structured comparison was conducted in the following stages: (1) Meetings and workshops between authors (i.e. expert group) from all four countries to understand the SML in each country; (2) deciding on questions about the SML through an iterative process to describe similarities and differences between the systems; (3) answering the questions in each country using information from sources such as experts, grey papers, unpublished materials, web pages and scientific papers. The answers were discussed among the researchers and checked with other experts in each country.

3. Results

All four countries have had e-prescribing and electronic storing of prescriptions for many years (Table 1). Denmark and Finland have implemented their solutions for a national SML, and Norway and Sweden are working on the implementation of their solution. The goal of the SML is similar in all countries: having one updated and correct medication list for all patients accessible for all health professionals can reduce the prevalence of medication errors, save time for health care professionals and increase patient safety (8-12). However, the information the SML is based on differs between the countries. The SML in Denmark and Norway is based on medication orders while Finland and Sweden have lists based on prescriptions (Figure 1).

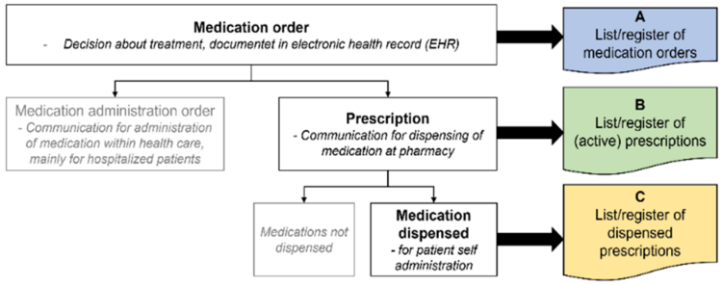


Figure 1. Overview of the different types of medication lists. The SML in Denmark and Norway is based on medication orders (A) but combined with information from active and dispensed prescriptions (B and C). The SML in Finland and Sweden are based on active prescriptions (B) combined with a record of dispensed prescriptions (C). Active/valid prescriptions are the basis for dispensing medications in all countries.

3.1. Shared medication list in Denmark

Denmark uses an electronic prescribing system called *Fælles Medicinkort* (FMK), a nationwide mandatory platform for communicating prescriptions. It is integrated with around 40 local solutions distributed in hospitals, general practices, municipalities, and others, meaning health professionals can see all medicines registered for every citizen directly through their local system (12). Physicians can also add, change and remove prescriptions in the FMK via their local systems or online. When a registration has been made, the medication will subsequently be synchronized with the central FMK. For inpatient care, the FMK is imported when the patient is admitted to the hospital, and the attending physician will manually transfer the hospital medication list to FMK when the patient is discharged (9). The FMK contains a list of all medication orders and prescription medications redeemed at the pharmacy within the last two years [9]. Over-the-counter (OTC) medications and dietary supplements that are not prescribed by a

physician, are not registered in the FMK [9]. Citizens have access to their FMK at www.sundhed.dk or the app "Medicinkortet", but cannot change the content [9]. The FMK was introduced in 2011 after four years of pilot testing. In 2014 it was fully implemented in general practice and hospitals, followed by a stepwise implementation up until 2019, when it was implemented in pharmacies.

3.2. Shared medication list in Finland

Finland has the Kanta services which are nationwide electronic healthcare services for healthcare professionals, community pharmacies, and citizens [13]. Kanta contains a centralized database called the Prescription Centre (ePC), where all issued prescriptions (both active and expired) and dispensing records are stored. The physician's EHR are integrated with the ePC so that all issued e-prescriptions are sent to the ePC [8]. The prescriber can also retrieve the patient's prescriptions from the ePC to view, renew, cancel or make changes to them. Pharmacy data systems search for prescriptions in the ePC, and the pharmacists can dispense them, request prescription renewal, and cancel and make corrections in some situations. Citizens have access to their prescriptions through the portal My Kanta [13]. The ePC has been introduced stepwise since 2010, and since 2017 it has been obligatory for all healthcare providers to use [13]. Even though all prescriptions are stored in the ePC, the information may not be up to date, e.g., the ePC may include prescriptions for medicines that the patient no longer takes or prescriptions that has undergone modifications, but where the changes have not been entered into the ePC. Therefore, Finland will introduce a national up-to-date Kanta medication list by the end of 2025 [8], including information on dose adjustments and discontinuations of medication. During 2027−2030 the Kanta medication list will also cover e.g. hospital medication and OTC medicines.

3.3. Shared medication list in Norway

Norway has the Prescription Intermediary, a database accessible to all prescribers and pharmacies in the country. The Prescription Intermediary contains all active prescriptions, but does not necessarily give a complete overview of the prescribed medications of a patient. For example, when a prescription expires (after 1 year) it will no longer be visible in this list, or it might contain outdated prescriptions if physicians do not cancel old prescriptions appropriately. Since 2012 a shared medication list, *pasientens legemiddelliste* (PLL) transmitted via the Prescription Intermediary has been planned. The current version of the PLL is only an overview of the current treatment and is not legally a prescription that can be used to dispense medicines, meaning that there must be e-prescriptions for each item on the PLL in addition to the PLL [11]. Other health care professionals can view, but not edit, the PLL in a portal called the Summary Care Record. Citizens have access to see their PLL through the national portal Helsenorge.no. The PLL system has been pilot tested on patients in home care receiving multidose drug dispensing in 2014 and 2018. Testing of PLL on patients with ordinary prescriptions started in 2021, and a nationwide implementation is planned to start in 2024 [11].

3.4. Shared medication list in Sweden

In Sweden e-prescriptions have been stored electronically in the National prescription repository (NPR) since 2005. Information about valid prescriptions in NPR has been available for pharmacists and patients, but not prescribers, resulting in prescribers

lacking information about prescriptions from other healthcare providers. With a new law in May 2021, the NRP was replaced by an SML called Nationella läkemedelslistan (NLL). The NLL includes valid prescriptions and information about dispensed prescriptions, and health care professionals are now allowed access [10]. The NLL is not yet integrated with healthcare EHR, but all prescriptions are transferred to NLL. The prescriber can use a web-based application called "Förskrivningskollen" to view NLL, prescribe and make changes to prescriptions [10]. However, any change must also be documented in the local EHR. All medication orders should be handled in the EHR, but only changes resulting in new prescriptions are transferred (i.e., not all instances of changed doses or termination of treatment), which can cause discrepancies between the prescriptions in the NLL and the medications the patients should take [1].

Table 1. Summary of the shared medication lists (SML) in Denmark, Finland, Norway and Sweden

Question	Denmark	Finland	Norway	Sweden
E-prescriptions	≥99%	≥99%	≥97%	≥99%
Electronic storing of prescriptions?	Since 2006	Since 2010	Since 2013	Since 2005
Name of the SML	Fælles medicinkort (FMK)	Prescription Service (ePS)	Patientens legemiddelliste (PLL)	Nationella läkemedelslistan (NLL)
Type of information in SML(based on Figure 1)?	A, B and C	B and C	A and B	B and C
When the SML was/will be implemented?	In stages since 2014	In stages from 2010.	Pilot testing since 2021	In stages since 2021.
Can the SML be seen directly in EHR?	Yes	Yes	Yes*	No, not yet***
Are medication changes in the EHR automatically changed in the SML?	Yes	No	Yes*	No ***
Mandatory and available in all sectors and regions?	Yes	Yes	Yes**	Yes ***
Access for citizens?	Yes	Yes	Yes	Yes
Can patients edit the SML?	No	No	No	No
Can the SML be used to administer medications for hospitalized patients?	No	No	No	No

*EHR: electronic health record. *only for those piloting the system as the system is not yet fully implemented. **Currently only available in one health region. ***the SML is planned to be integrated with the EHR and be mandatory to use, but details about functionality after integration are not clear*

The preliminary date for when all systems should finish the integration with the NLL is December 2025. However, it has not been clearly defined how NLL should be integrated with EHR and what fully implemented means. Citizens can access NLL using the web portal "Läkemedelskollen" or the patient portal 1177 "Journalen på nätet". Because NLL is based on prescriptions and the responsibility for ensuring the information is correct is not clearly defined, it is not intended to be used as a medication list for patients [10].

4. Discussion and Conclusion

Denmark, Finland, Norway and Sweden, are all in different stages of implementing SML systems, the systems are similar but there are also some differences. SML systems has potential to increase access to information and reducing discrepancies between lists, however, the information in the SMLs is still not always up to date (4-6). One reason for

discrepancies is that clinicians do not always update the SML appropriately (3). This can be linked to the integration with EHRs not supporting appropriate working routines due to legal or technical aspects or be related to human factors. Implementing SML systems raises the discussion about responsibility for updating the medication (12). If clinicians do not review the entire medication treatment or are reluctant to delete prescriptions by other physicians this might decrease medication safety (1, 4). In all four countries, citizens can view their SML, but not modify the list or give feedback to prescribers on the actual use of medications. Increased patient involvement will be necessary for having one accurate and complete medication list (2). In this paper, we describe four different solutions which are all referred to as a national SML, and in the literature, other kinds of solutions for a shared medication list are described with different words (7). Using a common terminology to describe the SMLs internationally will ease the comparison of the effects of such systems in the future. Future studies should further develop a thesaurus in this field, explore how new advanced technology could contribute, and investigate the patient perspective and possible privacy issues. All the Nordic countries have an SML solution, but none gives complete, up-to-date information about the medication use during hospital stays. Though citizens have access to view their SML in all countries, none has solutions where the citizens can take an active part, edit or stating their adherence in the SML. Although SMLs do seem to increase access to information about patient's prescribed medications and have the potential to reduce discrepancies, the knowledge about any effects of such system is still limited.

References

[1] Hammar T, Mzil L, Eiermann B. Discrepancies in patients' medication lists from pharmacies in Sweden: an interview study before the implementation of the Swedish National Medication List. Int J Clin Pharm. 2022:1-9.

[2] Sääskilahti M, et al. Benefits, Problems, and Potential Improvements in a Nationwide Patient Portal: Cross-sectional Survey of Pharmacy Customers' Experiences. J Med Internet Res. 2021;23(11):e31483.

[3] Bulow C, et al. Causes of discrepancies between medications listed in the national electronic prescribing system and patients' actual use of medications. Basic and Clinical Pharmacology and Toxicology. 2021;129:221-31.

[4] Josendal AV, et al. Implementation of a shared medication list in primary care - a controlled pre-post study of medication discrepancies. BMC health services research. 2021;21(1):1335.

[5] Bulow C,et al. Discrepancies Between the Medication List in Electronic Prescribing Systems and Patients' Actual Use of Medicines. Senior Care Pharmacist. 2019;34(5):317-24.

[6] Kauppinen H, Ahonen R, Mäntyselkä P, Timonen J. Medication safety and the usability of electronic prescribing as perceived by physicians-A semistructured interview among primary health care physicians in Finland. Journal of evaluation in clinical practice. 2017;23(6):1187-94.

[7] Manskow US, et al. Digital solutions for a shared medication list A narrative literature review. SHI 2019: November 12-13, 2019; Oslo, Norway: Linköping University Electronic Press; 2019.

[8] Kanta. Kanta Services 2022 [Available from: https://www.kanta.fi/en/.

[9] The Danish Health Data Authority. Fælles Medicinkort (FMK) (in Danish) 2022 [Available from: https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-faelles-medicinkort.

[10] Swedish eHealth Agency. Nationella läkemedelslistan [in Swedish] 2022 [Available from: https://www.ehalsomyndigheten.se/yrkesverksam/nationella-lakemedelslistan/.

[11] The Norwegian Directorate of e-health. Program pasientens legemiddelliste 2022 [03.03.22]. Available from: https://www.ehelse.no/programmer/program-pasientens-legemiddelliste.

[12] Pedersen L. Fælles medicinkort. In: Fredskild T, editor. Sundhedsinformatik i sundhedsvæsnet Gads forlag; 2018.

[13] Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. Finnish Journal of eHealth and eWelfare. 2018;10(4):381–95.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230105

An Architecture for Providing Personalized Digital Health

Jaime DELGADO [1] and Silvia LLORENTE
Departament d'Arquitectura de Computadors (DAC), Universitat Politècnica de Catalunya (UPC), Spain
ORCiD ID: Jaime DELGADO https://orcid.org/0000-0003-1366-663X
Silvia LLORENTE https://orcid.org/0000-0003-2000-6912

Abstract. Patients need mechanisms to integrate health information coming from different sources, including personal devices. This would lead to Personalized Digital Health (PDH). HIPAMS (Health Information Protection And Management System) is a modular and interoperable secure architecture that helps in achieving this objective and building a Framework for PDH. The paper presents HIPAMS and how it supports PDH.

Keywords. Personalized Digital Health, Modular architecture, Security, Privacy

1. Introduction

Integrating different sources of medical information into one system is one step towards achieving personalized health services. For this purpose, we can combine electronic health records coming from different health institutions with information coming from personal health devices, like smart watches or blood pressure monitors.

In order to provide a solution to facilitate this integration, we propose the use of a generic modular architecture called HIPAMS (Health Information Protection And Management System) [1] that is based on REST (REpresentational State Transfer) [2] web services. It is derived from work previously developed by the authors [3][4]. HIPAMS provides new specific modules to support medical information, like provenance and consents, as well as different modules for supporting different health information formats. There are other approaches for integration of different eHealth services, like the Profiles defined by IHE [5], where some of the concepts, like patient privacy consents, have already been defined.

We plan to propose HIPAMS as a way to develop a Personalized Digital Health Framework (PDH-F). In this context, ISO/TC 215/WG 11 (Personalized digital health) [6] has started to specify such a standard [7]. This PDH Framework specification will define how patients may manage their personal medical information in cooperation with health institutions.

In the rest of the paper, we describe HIPAMS and its modules and provide some hints on how such a platform can help in achieving the objectives defined in PDH-F.

[1] Corresponding Author: Jaime Delgado, E-mail: jaime.delgado@upc.edu.

2. Methods

The ideas behind HIPAMS come from existing work already developed for the secure management and protection of different kinds of digital information, like Multimedia content [3] and Genomic information [4]. Some initial ideas on the definition of HIPAMS were already presented in [1].

Figure 1 depicts its structure. The functionality of the most relevant modules is briefly described next. They are ordered alphabetically:

- Authorization Service: Module for access authorization based on privacy and access rules, and for validation of medical consents.
- Certification Authority: Provides digital certificates to secure communications.
- Consent Service: Module in charge of the creation and management of the medical consents.
- Health Content Service: Module in charge of health information management, both in reading and writing operations. This is a critical module covering the formats of the medical information, including Electronic Health Records (EHR) and medical documents, such as those based on HL7's CDA (Clinical Document Architecture) [8], among others.
- Metadata Service: Module in charge of handling metadata of medical documents, which may help in identifying and finding them. Again, very critical for the formats. In addition, FAIR (Findable, Accessible, Interoperable, Reusable) [9] [10] data principles could be considered.
- Policy Service: Module in charge of the creation of the authorization rules, which are organized into policies.
- Protection Service: Module which creates protection information as well as application of the mechanisms defined (i. e. encryption, signature, etc.).
- Provenance Service: Module in charge of adding and managing provenance information for health digital content supported by the platform.
- Report / Track Service: Module in charge of reporting the operations done in the system, especially those not authorized.
- Search Service: Performs searches over medical information, providing extra filtering features.
- User Application: Web application that sends requests to the Workflow Manager based on user actions. The communication between this application and the rest of the architecture is done through a secure channel.
- Workflow Manager: Intermediate module that acts as a unique entry point to the system. It checks operation authorization before interacting with other modules.

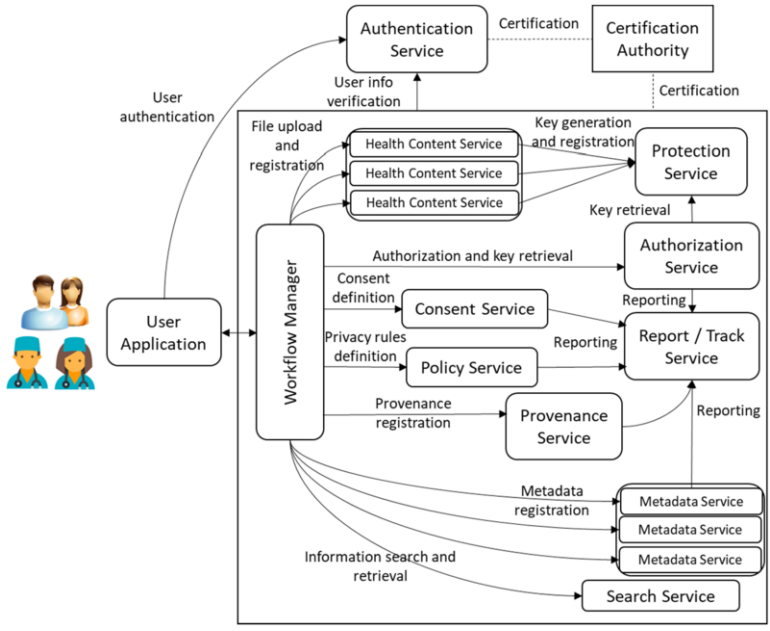


Figure 1. HIPAMS Architecture.

It is worth noting that this architecture allows the use of different formats representing medical information. The services managing these formats are Health Content and Metadata. As it can be seen in Figure 1, we may have several submodules inside Health Content and Metadata to support the different formats supported by the system implementing this architecture. Another approach for HIPAMS modules implementation could be using FHIR resources, such as Consents, or FHIR security aspects [11].

3. Results

We have defined several use cases to illustrate for example how HIPAMS can be used to register and search for medical information in a secure way. Figure 2 shows the workflow diagram for a registration use case, when the medical content requires encryption and privacy protection. The registered medical content may come from different sources (medical records, patient devices, …).

First of all, the user has to authenticate in front of the Authentication Service in order to receive a valid token. This token will be used afterwards when invoking the rest of services through the user application. Four services are used in this use case, Health Content Service (HCS), Metadata Service (MS), Protection Service (PtS) and Policy Service (PS). HCS is in charge of content creation. As protection is required, HCS calls the PtS corresponding operation to encrypt the medical information stored. Then, metadata is registered in MS. In this example, metadata does not need to be protected, but it might happen in other cases, so the PtS would be also called in that case. Afterwards, two privacy policies are created by the PS, to control later access to the protected medical

content. Again, the policies are created to control access to the content, but, if metadata also requires access control, one or more policies could be also created to control it.

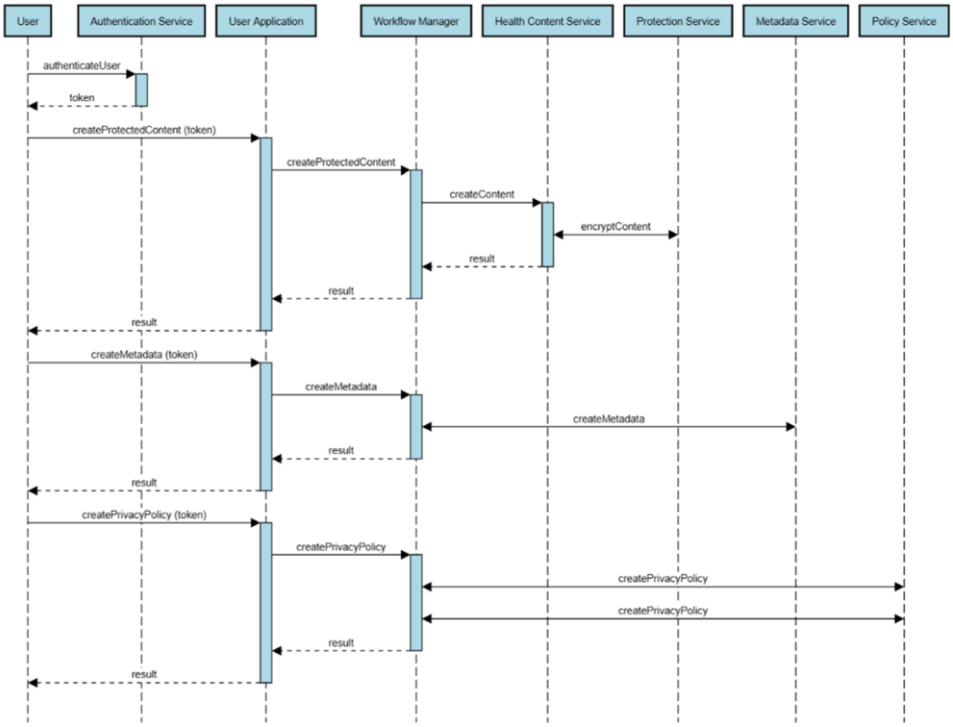


Figure 2. Protected medical content use case.

4. Discussion

HIPAMS is presented here as a possible way of facilitating Personalized Digital Health to patients. The final objective is to give them the opportunity to combine medical information coming from many different sources, like medical records coming from medical institutions and measures taken from their personal medical devices, like smart watches, mobile phones or any other medical device intended for being used at patients' home. This information may be useful for both primary and secondary use.

To do so, different medical information formats should be supported, making use of standardized formats as much as possible. As already shown in Figure 1, HCS and MS may contain submodules to support different formats. Moreover, if new formats appear in the future, they could be supported by implementing / integrating new submodules.

We would like to highlight that the architecture allows the inclusion of externally implemented modules, by defining a REST Application Programming Interface (API) that then calls to the module.

Providing this additional information to healthcare professionals facilitates a better medical decision, but may imply the need of extra work for its processing. However, HIPAMS provides tools, such as advanced search, that reduce this extra burden.

5. Conclusions and Future Work

We have introduced HIPAMS as an architecture to facilitate sharing and integration of health information coming from different sources with a special focus in patients' personal information. This leads to PDH (Personalized Digital Health).

Nevertheless, some specific aspects need further development, such as the idea of dynamic consents, which is implicit in the model, by jointly using Consent and Authorization services. Another example is the Provenance service, which is also very relevant.

On the other hand, HIPAMS could be used for the development of the PDH Framework to be standardized by ISO/TC 215/WG 11 "Personalized digital health". We are currently contributing to this work. With this purpose in mind, APIs should be defined for the different modules and more use cases identified.

Finally, and going further in the issues raised in the Discussion section, the concept of a federation of HIPAMS could be considered in order to support different medical institutions, as it is not realistic to store all this information in only one platform. The definition of standardized interfaces is key to achieve this objective. An advantage of having such a federation is that the communication between equivalent modules could be done through the same operations.

Acknowledgements

The work presented in this paper has been partially supported by the Spanish Government under the project GenClinLab-Sec (Mechanisms for secure and efficient management of genomic information tailored to clinical laboratories: Security Aspects, PID2020-114394RB-C31) funded by MCIN/AEI/10.13039/501100011033 and by the Generalitat de Catalunya (2017 SGR 1749).

References

[1] Delgado, J, Llorente, S. Provenance and dynamic consents for the management of medical data. Studies in Health Technology and Informatics, 2022, Volume 299: 171-176. doi: 10.3233/SHTI220978
[2] Fielding RT. Architectural Styles and the Design of Network-based Software Architectures [PhD. Dissertation]. University of California, Irvine; 2000
[3] Llorente S, Rodriguez E, Delgado J, Torres-Padrosa V. Standards-based architectures for content management. IEEE multimedia. 2013 Oct.-Dec.; 20(4):62-72. doi: 10.1109/MMUL.2012.58
[4] Llorente S, Delgado J. Implementation of Privacy and Security for a Genomic Information System Based on Standards. Journal of Personalized Medicine. 2022; 12(6):915. doi: 10.3390/jpm12060915
[5] IHE International. Integrating the Healthcare Enterprise. Available from: https://www.ihe.net/
[6] ISO/TC 215. Health informatics Information. Available from: https://www.iso.org/committee/54960.html
[7] ISO/TC 215/WG11, ISO/AWI TS 6201, Health Informatics - Personalized Digital Health Framework (PDH-F) (Under development). Information available from: https://www.iso.org/standard/82107.html
[8] HL7. HL7 Clinical Document Architecture (CDA). Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7
[9] Wilkinson MD, Dumontier M, Aalbersberg IJ et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3. 2016; 160018. doi: 10.1038/sdata.2016.18
[10] Delgado J, Llorente S. FAIR Aspects of a Health Information Protection and Management System. Methods of Information in Medicine. 2022 Dec;61(S 02):e172-e182.
[11] HL7. HL7 Fast Healthcare Interoperability Resource (FHIR). Available from: https://hl7.org/fhir/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230106

Social Determinants of Health Data Quality at Different Levels of Geographic Detail

Melody L. GREER[a,1], Maryam Y. GARZA[a], Steve SAMPLE[a] and Sudeepa BHATTACHARYYA[b]
[a] *University of Arkansas for Medical Sciences, USA*
[b] *Arkansas State University, USA*
ORCiD ID: Melody L. Greer https://orcid.org/0000-0002-1070-5907
ORCiD ID: Maryam Y. Garza https://orcid.org/0000-0002-2652-5935
ORCiD ID: Steven Sample https://orcid.org/0000-0003-1544-0458
ORCiD ID: Sudeepa Bhattacharyya https://orcid.org/0000-0003-3493-6029

Abstract. Social determinants of health (SDOH) impact 80% of health outcomes from acute to chronic disorders, and attempts are underway to provide these data elements to clinicians. It is, however, difficult to collect SDOH data through (1) surveys, which provide inconsistent and incomplete data, or (2) aggregates at the neighborhood level. Data from these sources is not sufficiently accurate, complete, and up-to-date. To demonstrate this, we have compared the Area Deprivation Index (ADI) to purchased commercial consumer data at the individual-household level. The ADI is composed of income, education, employment, and housing quality information. Although this index does a good job of representing populations, it is not adequate to describe individuals, especially in a healthcare context. Aggregate measures are, by definition, not sufficiently granular to describe each individual within the population they represent and may result in biased or imprecise data when simply assigned to the individual. Moreover, this problem is generalizable to any community-level element, not just ADI, in so far as they are an aggregate of the individual community members.

Keywords. Data Quality, Social Determinants of Health

1. Introduction

Social determinants of health (SDOH) impact 80% of health outcomes [1] from acute to chronic disorders. Interest in this information has been increasing over the last decade and the Centers for Medicare & Medicaid Services (CMS) Comprehensive Primary Care Plus (CPC+) Model requires providers to assess patients' social risks to aid in more accurate care delivery [2]. Attempts are underway to consistently provide these data elements to researchers and clinicians. It is, however, difficult to collect SDOH data as the main mechanisms used are (1) surveys which require time for documentation, and are not standardized across institutions, [3] or (2) aggregates at the neighborhood level [4]. In order to provide value, SDOH data must be complete, current, accurate. But neighborhood level data elements are not specific to each individual and so may not be correct for every patient. To study this relationship and understand the level of error that

[1] Corresponding Author: Melody L. Greer, E-mail: mlgreer@uams.edu.

results from using neighborhood level data as a stand-in for individual level data we compared the Area Deprivation Index to individual and their household level social determinants.

The ADI is built from census data and includes the factors of income, education, employment, and housing quality [5]. The American Community Survey five-year estimates are used to produce the ADI which is a neighborhood level index available freely for download at the state or national level in both census blocks and ZIP Code versions. This validated index has been available for 30 years and has seen increased usage over the last several years. The ADI does a good job of representing populations [6]. It does not, however, purport to represent individuals and should therefore not be used as a replacement for individual level representations. Aggregate measures do not describe each individual within the population they represent. For example, if certain SDOH elements like education or income are highly variable within a neighborhood, the neighborhood-level aggregate will be an imprecise estimate of any individual's actual measure. This idea is generalizable to any community-level data element as they are an aggregate of the measures of individual community members.

2. Methods

We created an integrated data set composed of approximately 55,000 electronic health records linked to the state level ADI and commercial consumer data. The ADI was obtained by downloading the census block group level ADI score and the zip+4 centroids contained within each block group. In previous work we have developed a repeatable SDOH enrichment and integration process to incorporate dynamically evolving SDOH domain concepts from consumers into clinical data. [7] In the course of that work we demonstrated that commercial consumer data can be a viable source of SDOH factors at an individual-level for clinical data providing a path for clinicians to improve patient treatment and care. [8, 9]

Because ADI is composed of income, education, employment, and housing quality data we have chosen these elements from the commercial consumer data for comparison. We computed the average ADI within each zip code. Commercial consumer data elements were not available for every individual or their household in the study, and those with null values were eliminated from the calculations, but were retained for other data quality comparisons.

3. Results

As shown in Figures 1-4 the box plots reflect the variance of the summarized ADI within the different levels of each demographic element and demonstrates the potential for inaccuracy when applying neighborhood level data to an individual. Although the neighborhood level elements skew as expected the overlap in actual values with respect to disparity is problematic. This overlap is a result of the wide variance for each of the categories.

In addition to accuracy, it is important to consider other aspects of data quality for use in patient care, such as completeness and timeliness. Upon review of both datasets, commercial consumer data vs. ADI, there were clear differences with regards to data completion. Typically, commercial consumer data is available for many individuals, but

may not be complete across all factors or data elements (e.g., for people who use only cash, have no subscriptions, and do not use discount cards). The commercial consumer data we used in this analysis was 98.98% complete for income, 62.63% complete for education, 48.23% complete for employment, and 94.69% complete for housing quality across all individuals in the dataset. In comparison, because the ADI is an aggregate measure, data for all factors was 'complete' for every individual or their household.

In regards to timeliness, the commercial consumer data was no more than six months old when purchased whereas ADI, and other indices constructed using census data, can only be as current as the last five-year estimate. Because commercial consumer data is purchased for an agreed-upon amount of time that can include a regular refresh schedule, it is always current.

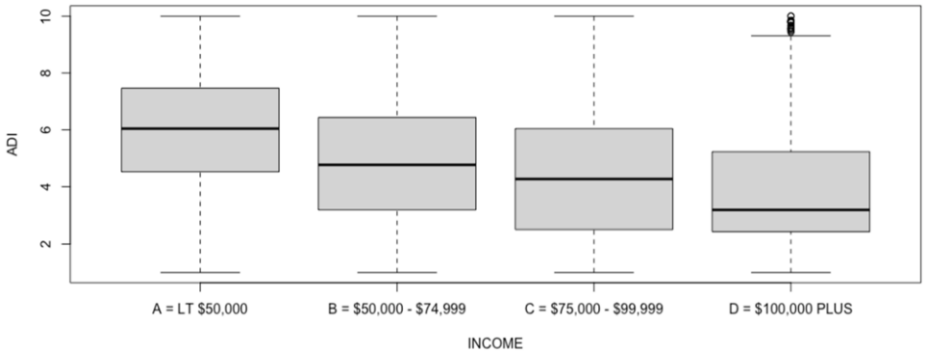


Figure 1. Income value compared with ADI.

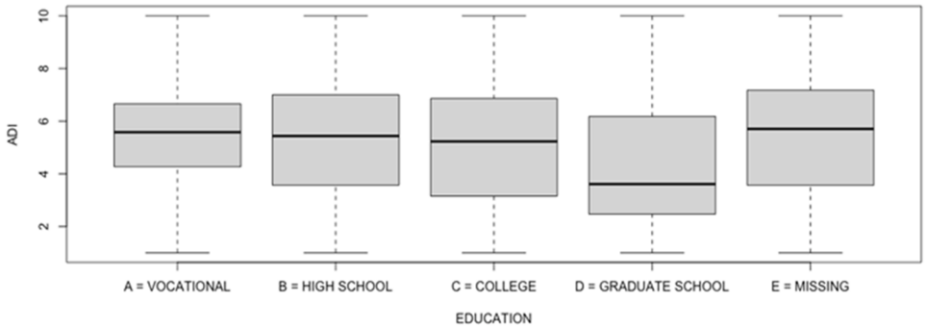


Figure 2. Education level compared with ADI.

Figure 3. Occupation value compared with ADI.

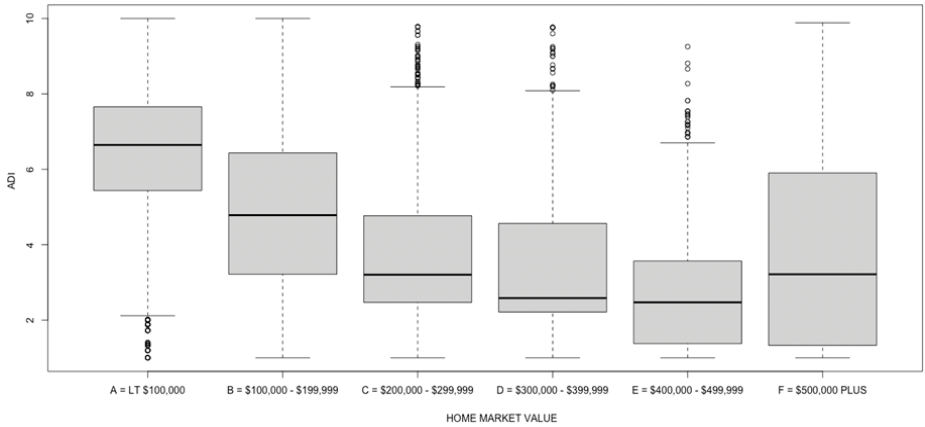


Figure 4. Home market value compared with ADI.

4. Discussion

This work demonstrates that neighborhood level SDOH elements are not sufficiently granular to provide clinicians or researchers individual level information. Other researchers have conducted related research with different deprivation indices and have come to similar conclusions. [4] Further, research suggests that the relationship between health outcomes and neighborhood level social determinants of health are not caused by the neighborhood environment but result from sorting by economic means [10].

4.1. Accuracy

The box plots reflect the range of ADI values within each category, revealing the broad range of deprivation for households and individuals. This demonstrates that ADI is associated on a large geographic scale but not on a smaller one. The standard deviation and variance of income, education, employment, and housing quality shown in Figures 1-4 are clear indications that these population level elements would often be inaccurate if used for individual patients in medical decision-making. Doing so opens the door for misclassification of patient risk level and invalidates the output of any would be automated clinical decision support systems.

4.2. Completeness

Missing information can affect the quality of care a patient receives. This problem is not as serious as wrong information, but can still have a negative impact on medical decision-making, especially if the problem is large and pervasive. Because ADI is an aggregate value, completeness is not an issue. However, a weakness of commercial consumer data is that it is not always available for every individual or their household. In this work, we have ignored null values in an effort to focus on accuracy, but this cannot be a solution for clinical applications.

4.3. Timeliness

For healthcare purposes, it is important to have current social context information. Frequently, individuals will change habits, move, and switch employers. These changes

directly impact factors related to their health in both positive and negative ways. If the information provided to clinicians or clinical decision support systems is out of date, this could have a negative impact on patients by delaying access to care. Commercial consumer data can be constantly refreshed to maintain the optimal current status. As far as we are aware, this is not true for publicly available data, and it is definitely not the case for any elements that use census data.

5. Conclusions

Because social context information is valuable in clinical practice and research, SDOH data elements have received an increasing amount of attention in the last two decades. Consistently collecting this data in a standard format remains a problem. Clinical providers are extremely busy, and additional data entry is not an optimal solution. In addition, SDOH screening instruments are not standardized, making interoperability challenging, if not impossible. To address this difficulty, social data aggregated at the neighborhood level has been proposed as a surrogate. However, this solution is also problematic as we have seen with the ADI because individual level data is not equivalent to neighborhood level data in every case. Moreover, aggregate measures are, by definition, not on target for each individual within the population which they represent. As a result, we conclude that data used to understand the social context for healthcare should be complete, current, accurate, and specific to the individual and their unique household.

References

[1] Institute for Clinical Systems Improvement. Going beyond clinical walls: solving complex problems. 2014.
[2] Centers for Medicare & Medicaid Services. Comprehensive Primary Care Plus CMS.gov2022 [cited 2022 12/15]. Available from: https://innovation.cms.gov/innovation-models/comprehensive-primary-care-plus.
[3] Cottrell EK, Dambrun K, Cowburn S, Mossman N, Bunce AE, Marino M, et al. Variation in electronic health record documentation of social determinants of health across a national network of community health centers. American journal of preventive medicine. 2019;57(6):S65-S73.
[4] Cottrell EK, Hendricks M, Dambrun K, Cowburn S, Pantell M, Gold R, et al. Comparison of community-level and patient-level social risk data in a network of community health centers. JAMA network open. 2020;3(10):e2016852-e.
[5] Health UoWSoMaP. Area Deprivation Index. In: Research CfHD, editor. 2020.
[6] Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J. Introduction of an area deprivation index measuring patient socioeconomic status in an integrated health system: implications for population health. EGEMs. 2016;4(3).
[7] Greer ML, Zayas CE, Bhattacharyya S. Repeatable enhancement of healthcare data with social determinants of health. Frontiers in big Data. 2022;5.
[8] Greer ML, Sample S, Jensen HK, McBain S, Lipschitz R, Sexton KW. COVID-19 Is Connected with Lower Health Literacy in Rural Areas. Studies in Health Technology and Informatics. 2021;281:804-8.
[9] Wong L, Yu F, Bhattacharyya S, Greer ML. Covid-19 Positivity Differences Among Patients of a Rural, Southern US State Hospital System Based on Population Density, Rural-Urban Classification, and Area Deprivation Index. Studies in Health Technology and Informatics. 2022;294:701-2.
[10] Jokela M. Are neighborhood health associations causal? A 10-year prospective cohort study with repeated measurements. American journal of epidemiology. 2014;180(8):776-84.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230107

Development of OSOMO Prompt Mobile Application on Elderly Population for Village Health Volunteers Using the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) Model

Niruwan TURNBULL[a,b], Kukiat TUDPOR[a,b], Siwaporn CHIMPHOKLANG[a,b], Sureewan SILADLAO[c], Nutchanun KAMONROEK[d], and Kannikar Hannah TATIYAWORAWATTANAKUL[e,1]
[a] *Public Health and Environmental Policy in Southeast Asia Research Cluster (PHEP-SEA)*
[b] *Faculty of Public Health, Mahasarakham University*
[c] *Suan Sunandha, Rajabhat university*
[d] *Ubon Ratchathani Provincial Health Office*
[e] *Public Health Department, Torrens University Australia*
ORCiD ID: Niruwan TURNBULL: https://orcid.org/0000-0002-7698-3352
Kukiat TUDPOR: https://orcid.org/0000-0002-8533-5891
Siwaporn CHIMPHOKLANG: https://orcid.org/0000-0002-3208-0813
Sureewan SILADLAO: https://orcid.org/0000-0002-4162-1335
Nutchanun KAMONROEK: https:// 0000-0002-7616-6213
Kannikar Hannah TATIYAWORAWATTANAKUL: https://orcid.org/0000-0002-2901-1606

Abstract. This paper aims to describe the use ADDIE model in developing a digital health tool, OSOMO Prompt app, and discuss evaluation outcomes of using this digital tool by village health volunteers (VHV) in rural areas in Thailand. The OSOMO prompt app was developed and implemented in elderly populations in eight rural areas. The Technology Acceptance Model (TAM) was used to test the acceptance of the app four month after the implementation. There were 601 VHVs voluntarily involved in the evaluation phase. The ADDIE model was successfully employed to guide the research team to develop the OSOMO Prompt app consisting of four services delivered to elderly populations by VHVs, including: 1) health assessment; 2) home visit; 3) knowledge management; and 4) emergency report. The findings from the evaluation phase reported that the OSOMO Prompt app was accepted as utility and simplicity (score 3.95+.62); and valuable digital tool (score 3.97+.68). The app received the highest score for being a useful tool assisting VHVs in achieving their work goals and improving work performance (score 4.0+.66). The OSOMO Prompt app could be modified for other healthcare services in different

[1] Corresponding Author: Kannikar Hannah TATIYAWORAWATTANAKUL, Public Health Department, Torrens University Australia; 88 Wakefield Street, Adelaide SA 5000; E-mail: hannah.wechkunanukul@torrens.edu.au

populations. Further investigation in long-term use and its impact on healthcare system is warranted.

Keywords. ADDIE model, Mobile Application, Digital Health Technology, Elderly, Village Health Volunteer

1. Introduction

Digital transformation has been implemented into healthcare system globally to improve health provisions, strategies, and operations through digital technology. Different types of digital tools have been developed and used in many health organisations such as mobile applications, computerized-based data system, and tracing system. This digitalization has enhanced a quality of health data in various aspects, including the timely and accurate information, optimizing data collection, and efficient data analysis which led to effectiveness and sustainability of healthcare practices (1,2). Mobile applications have been considered as a beneficial technology integrated in epidemiological research to cope with the challenges in data collection process (3).

In Thailand, health-related data in communities have been collected manually by the health volunteers on paper based practices for many decades which has posed significant challenges to quality of data collection and also burden of work. There is limited evidence of using digital tools in health data collection in rural area in Thailand. To improve the data collection and data analysis procedure at community level, the specific mobile application has been developed for community healthcare workers. This paper aims to describe the development of mobile application for village health volunteers (VHV) using the ADDIE model and discuss evaluation outcomes of this digital tool at a community level in Thailand.

2. Methods

OSOMO prompt mobile application is a digital tool developed for health volunteer workers to collect health data of elderly populations in rural communities. The Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model was used to guide the development of this mobile app. This research project has been approved by the Maha Sarakham Provincial Public Health Office's Ethical Review Committee for Human Research, chairperson: Mr. Pakee Sappipat (No. 6/2564).

2.1. Analysis phase

The survey was conducted to gain insight into socioeconomic and health status of elderly group, including nine domains: 1) Socioeconomic status, 2) Heath status, 3) Stress levels, 4) Quality of life,5) Activities of Daily Living (ADL), 6) Health knowledge, 7) Attitude toward health, 8) Health literacy, and 9) Health behavior.

2.2. Design and Development Phase

The content was designed upon on data analysis and using geographic information systems on smartphones and were transformed to a hybrid application, including a web-

based app and mobile apps for iOS and Android. The smartphone application has been entitled 'OSOMO Prompt' (Figure 2). There were 8,348 of aged populations and 1,019 VHV involved in this phase covering eight areas of four provinces in the Northeast region of Thailand (Roi Et, Khon Kaen, Maha Sarakham, and Kalasin) (Figure 1).

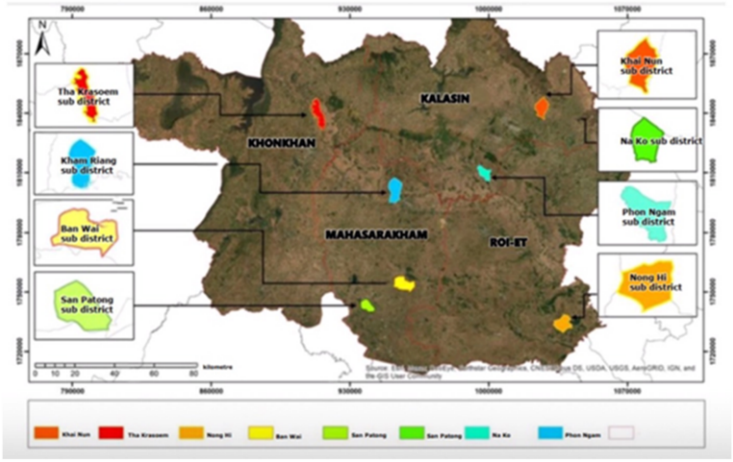


Figure 1. Map of the eight study areas in the Northeast region of Thailand

Figure 2. Responsibility diagram and the fetched information of OSOMO app

2.3. Implementation Phase

A pilot study deployed a quasi-experimental study design (before-after trial) were conducted in eight areas, including 30 VHVs selected from each area. In total, 240 VHVs were recruited in this phase. The participants were asked to use the OSOMO Prompt app to help manage their routine tasks and collect health data in four domains: 1) health assessment, 2) home visit, 3) knowledge management, and 4) emergency report. The pilot study revealed the efficacy of the OSOMO Prompt app and the results have been published elsewhere (4). The app was then implemented in VHV practices in eight areas of the project.

2.4. Evaluation Phase

There were two stages of the evaluation process. The first stage was a preliminary evaluation occurred during the design and development phase before the implementation. It was a system evaluation performed by three experts to assess for appropriate language,

computer language, and accuracy. The second evaluation occurred four months after the implementation using the Technology Acceptance Model (TAM) to assess acceptability and feasibility of the app (5). Data were voluntarily collected from 601 VHVs who used this app and worked in the eight areas of the project. The full report of TAM has been published elsewhere (6).

3. Results

The OSOMO Prompt app was successfully developed and implemented using the ADDIE model. The main menu consists of four services delivered by VHVs to the elderly populations in rural areas: 1) health assessment; 2) home visit; 3) knowledge management; and 4) emergency report (Figure 3). Evaluation data were collected based on the two key aspects of the OSOMO prompt app. The findings derived from a pilot study revealed a significant increase in acceptance score of before and after using the OSOMO prompt (10.49+2.53 and 12.18+2.76 respectively $p<0.001$). Additionally, the results of evaluation phase revealed that majority of participants reported the OSOMO prompt app as utility and simplicity. The app was reported as easy to use tool (score 3.95$\pm$.62) and was accepted as a valuable tool in VHV work (score 3.97$\pm$.68). 'Using OSOMO prompt app makes it easy to accomplish VHV jobs' received the highest mean score at 4.0+.66.

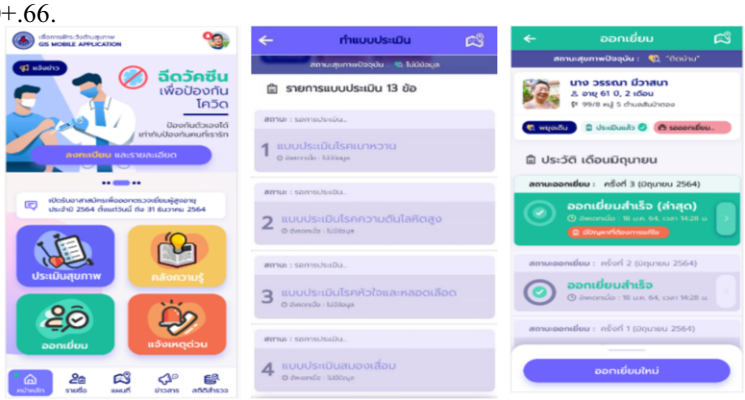


Figure 3. The main menus of OSOMO prompt mobile application.

4. Discussion

The results showed that the ADDIE model were successfully applied to guide the research team in developing a digital tool, the OSOMO Prompt app, for healthcare service and health data collection in rural areas. The ADDIE model was originally developed for a learning media and has been applied to several studies in care management such as Educare app, a mobile application for clinical duty of nursing students and nurse educators (7) and application for managing post-surgical symptoms for patients (8). The findings of evaluation phase revealed that VHVs perceived the OSOMO Prompt app as a helpful and useful tool for their work. This positively indicates the efficacy and acceptability of integrating digital innovation into healthcare practices in rural areas to improve quality of provision and increase professional development.

The high score in acceptance confirmed that digital technology is an acceptable tool among communities with low socioeconomic status and it is also been considered as a practical tool in relation to data collection and data analysis for healthcare system. The findings of this study provide evidence-based practice for healthcare provision delivered by health volunteers at community level in Thailand. Assessing sustainability and long-term outcomes of this app would be a plan for the next step. Additionally, it is warranted for a future study in using OSOMO Prompt app in different health services and different populations such as tracking pregnancy and prenatal care among unreported pregnancy in rural areas and monitoring self-management for chronic disease in ethnic groups.

5. Conclusion

The outcomes of this study demonstrated the achievement of using the ADDIE model to guide the development of the OSOMO Prompt mobile app for Village health volunteers to deliver healthcare services to elderly populations in rural areas in Thailand. The OSOMO Prompt app has been accepted as a utility and simplicity tool among the users and also assist VHVs to improve quality of their works and career development. The OSOMO Prompt app could be adapted and implemented in different healthcare services for different populations. Future research in long-term outcomes from using this digital tool is warranted.

Acknowledgements

This research was financially supported by Mahasarakham University.

References

[1] Aziz HA. A review of the role of public health informatics in healthcare. *Journal of Taibah University Medical Sciences*. 2017;12(1):78-81 [PMID:PMID DOI:https://doi.org/10.1016/j.jtumed.2016.08.011]
[2] Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. *Annu Rev Public Health*. 2015;36:345-59 [PMID:PMID DOI:https://doi.org/10.1146/annurev-publhealth-031914-122747]
[3] Fischer F, Kleen S. Possibilities, Problems, and Perspectives of Data Collection by Mobile Apps in Longitudinal Epidemiological Studies: Scoping Review. *J Med Internet Res*. 2021;23(1):e17691 [PMID:PMID DOI:10.2196/17691
[4] Turnbull N, Wongkongdech A, Wongkongdech R, Chaiyakarm T, Jitsukka W, Sombateyotha K, Leethongdeesakul S, Kuboonya-Aragsa N, Chantarsombat C, Kimchai K, Kamonroek N. Implementation of a Geographic Information System-Based Smartphone Application for Health Vigilance in Older Adults by Village Health Volunteers (OSOMO). Advances in Informatics, Management and Technology in Healthcare. 2022;295:242. https://doi.org/10.3233/shti220707
[5] Davis, F.D. and Venkatesh, V., 1996. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International journal of human-computer studies*, *45*(1), pp.19-45. https://doi.org/10.4018/978-1-4666-8156-9.ch013
[6] Turnbull N, et al. Acceptability and Feasibility of Smart Phone Application" OSOMO Prompt" for Health Surveillance in Older Adults by Community Health Workers. Studies in Health Technology and Informatics. 2022 Jun 1;295:386-9.
[7] Positos JD, Abellanosa ALA, Galgo CAL, Tecson CMB, Ridad GS, Tabigue MM. Educare App: Mobile Application for clinical duties of nursing students and nurse educators. Enfermería Clínica. 2020;30:12-
[8] AYDIN A, GURSOY A. Mobile app for breast cancer patients: Development process using the ADDIE models (Step 1). 2022.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230108

Implementation, Adoption and Use of the Kanta Services in Finland 2010–2022

Vesa JORMANAINEN[a,b,1], Tuulikki VEHKO[a] , Marina LINDGREN[c], Ilmo KESKIMÄKI[a,d] and Minna KAILA[b]
[a] *Finnish Institute for Health and Welfare (THL)*
[b] *University of Helsinki, Faculty of Medicine, Department of Public Health*
[c] *Social Insurance Institution of Finland (Kela)*
[d] *Tampere University, Faculty of Social Sciences*
ORCiD ID: Vesa Jormanainen https://orcid.org/0000-0002-1906-8043

Abstract. Nationwide implementation and adoption of the Prescription Centre and the Patient Data Repository services required 5.5 years since May 2010 in Finland. The Clinical Adoption Meta-Model (CAMM) was applied in the post-deployment assessment of the Kanta Services in its four dimensions (availability, use, behavior, clinical outcomes) over time. The CAMM results on the national level in this study suggest 'Adoption with Benefits' as the most appropriate CAMM archetype.

Keywords. Implementation, adoption, use, Kanta Services, Finland, Clinical Adoption Meta-Model

1. Introduction

Healthcare systems are currently facing several challenges, such as population aging, economic constraints and rapid technological change, which call for comprehensive reforms. Internationally a prevailing solution offered for these challenges is to deploy large-scale information and communication technology systems [1,2]. However, adopting new national health information (HIS) or health information exchange (HIE) systems are often challenging ventures: a large-scale, complex and costly endeavor, taking many years to develop and build, involving multiple public and private stakeholders and that has impact on millions of people [3–7].

HIS integrates the data collection, processing, reporting, and use of the information necessary for improving health service effectiveness and efficiency through better management at all levels of health services [8,9]. HIE essentially includes the electronic transfer of patient data and health information between healthcare service providers or institutions [10–13].

Current literature is still dominated by reports of single organizations and the punctuality of their HIS implementation [14–17]. Moreover, most HIS/HIE system and healthcare reforms are not properly followed up and their outcomes are rarely evaluated [2,17].

[1] Corresponding Author: Dr. Vesa Jormanainen, Finnish Institute for Health and Welfare, P.O. Box 30, 00271 Helsinki, Finland; E-mail: vesa.jormanainen[at]gov.fi.

This research aimed to assess the implementation and adoption of the *Act on Electronic Prescription* (61/2007) and the *Act on Processing Customer Data in Health and Social Care* (159/2007). The assessment utilized log-based register data on use of the national healthcare Kanta Services since May 2010 in Finland.

2. Methods

The Kanta Services is the name of Finland's centralized, shared, and integrated electronic data system services introduced in phases since May 20, 2010 [18–20]. The implementation processes and adoption efforts in healthcare were prospectively followed-up by utilizing records accumulated in the Finnish Institute for Health and Welfare's (THL) national operative coordinating unit from May 2010 to December 2016. Use of the Kanta Services from May 2010 to December 2022 was followed-up by rigorously utilizing log-based register data on descriptive performance indicators' time series provided by the Social Insurance Institution of Finland (Kela).

We used Clinical Adoption Meta-Model's (CAMM) four dimensions (availability, use, behavior, clinical outcomes) over the fifth dimension of time [21,22]. According to the CAMM, a data system will become available, it will be used which in turn will lead to changes in behavior, that will produce clinical or health outcomes over time.

In the 'No Deployment' archetype of the CAMM, the HIS fails to reach end-users. In the 'Low Adoption' archetype, the HIS is deployed and available, but with minimal or rapidly declining use. In the 'Adoption without Benefit' archetype, a HIS is deployed, available and used by end-users, but it fails to achieve the intended behavior changes or the expected outcomes. In turn, the 'Behavior Change without Outcome Benefit' archetype occurs when an adopted HIS produces the expected behavior change but fails to produce the expected outcomes. The 'Adoption with Benefits' archetype is characterized by a clear progression of HIS availability that leads to ongoing HIS use, which then causes observable changes in clinical and health behavior that, in turn, result in improvements in measured outcomes. In the 'Benefit without Use' archetype, the expected behavior changes and/or outcomes occur without HIS use.

3. Results

3.1. Availability of the nationwide healthcare Kanta Services

The first certificate of interoperability acceptance for healthcare and pharmacy data systems for the electronic Prescription Centre services was granted in May 18, 2010, and in public healthcare, it took 859 days to reach the 50% national population coverage and 1,258 days for the 100% point.

The first certificate of interoperability acceptance for healthcare data systems for the electronic Patient Data Repository was granted in October 28, 2013, and in public healthcare, it took 376 days to reach the 50% point and 760 days for the 100% point.

The mean implementation and adoption time varied considerably (Table 1).

Table 1. Mean number of years from the start of the first implementation and adoption project to the start of the last implementation and adoption project (min. and max. times in years in brackets) of the healthcare Kanta Services in Finland in 2010–2015. PHCs refers to public primary healthcare centres.

Kanta Services to be implemented	**Hospital districts** Mean (min.–max.)	**University-hospital-specific catchment areas** Mean (min.–max.)	**National** Mean
Prescription Centre			
Pharmacies, years	0.6 (0.1–0.2)	1.3 (0.9–2.0)	2.4
PHCs, years	0.6 (<0.01–2.0)	2.0 (0.9–3.4)	3.4
Pharmacies+PHCs, years	1.1 (0.02–2.4)	2.3 (1.8–3.4)	3.4
Patient Data Repository			
PHCs, years	0.6 (<0.01–1.6)	1.5 (1.1–2.0)	2.0
Both Kanta Services			
PHCs, years	3.0 (1.6–5.1)	4.3 (3.6–5.1)	5.5
Pharmacies+PHCs, years	3.6 (2.5–4.3)	4.6 (4.4–5.1)	5.5

3.2. Use of the national healthcare Kanta Services in Finland since May 2010

Citizen users have accessed the web based My Kanta Pages since May 2010, sent electronic repeat prescription (eP) renewal requests to healthcare organizations since November 2015, and have made visits on behalf of their children since October 2016 (Table 2). In addition, citizen users have recorded consents, consent restrictions, organ donation testaments and living wills into the Patient Data Management Service.

Table 2. Performance indicators by healthcare Kanta Services in 2022 and cumulatively 2010–2022 in Finland (numbers in millions). ePs refers to electronic prescriptions.

Kanta Service Performance indicator	**2022** Millions	**2010–2022** Millions
My Kanta Pages		
Sign-ins	37.2	179
Repeat eP renewal requests	3.4	18.7
Visits on behalf of children	2.5	12.2
Prescription Centre		
ePs	28.2	274
Dispensation documents	76.7	580
Patient Data Repository		
Documents	477	3174
Service events	230	1581
Persons	6.5	6.5
Information notices	0.24	7.9
Consents	0.24	4.8
Consent restrictions	0.12	0.16
Organ testaments	0.05	0.88
Living wills	0.04	0.28

Healthcare professionals (community pharmacy professionals since 2017) have recorded ePs, and community pharmacy professionals medication dispensing documents into the Prescription Centre since May 2010 (Table 2). Healthcare professionals have recorded documents related to service events into the Patient Data Repository since November 2013, and information notices into the Patient Data Management Service.

3.3. Behavior of healthcare professionals

Healthcare professionals quickly learned to issue and use ePs, and pharmaceutical professionals in community pharmacies learned quickly to record paper-based and

telephone prescriptions into the Prescription Centre after the 'big bang' change in January 2017, when ePs became mandatory in Finland.

The Kela launched free-of-charge Kelain web-based service in September 2016 to support the start of mandatory ePs in January 2017. The number of registered healthcare professional Kelain service users rose rapidly to 18,000 and the number of ePs issued via Kelain service rose to 0.301 million in 2018.

3.4. CAMM archetypes and the Kanta Services in Finland

The Kanta Services already reached end-users in a clinical setting in May 2010, and, thus, escaped the CAMM 'No Deployment' archetype. The Kanta Services' implementation and adoption matured and escaped the archetypes 'Low Adoption', 'Adoption without Benefit', 'Behavioral Change without Outcome Benefit' and the 'Benefit without Use'. Based on the results of this study, the most accurate CAMM archetype is 'Adoption with Benefits'.

4. Discussion and conclusions

The results of this research suggest that it is possible to implement and adopt two large-scale national HIS/HIE in 5.5 years covering all public primary healthcare centres, community pharmacies and hospitals together with most private healthcare service providers in a country with 5.5 million inhabitants.

An implementation strategy combining top-down and bottom-up approaches [2,11,23–26] employed in Finland proved an appropriate strategy for the implementation and adoption of the two healthcare Kanta Services.

The Prescription Centre services were implemented and adopted first and thereafter the Patient Data Repository services [18,19]. Public healthcare providers implemented and adopted the Kanta Services first and thereafter private healthcare service providers. The initial hospital district based implementation strategy for the Prescription Centre services informed the strategy to a certified Kanta Services compatible HIS/HIE strategy for the Patient Data Repository services. The THL's operative coordination unit – with appropriate legal mandate in the permanent legislation – supported the large-scale implementation processes of the two healthcare Kanta Services in close cooperation with the Kela.

The results from a previous study showed that the implementation of HIS/HIE systems in Finland has successfully passed several milestones in terms of the CAMM archetypes [27]. The case of Kanta Services in Finland suggested that the healthcare-specific CAMM theoretical construct can be used on the national level HIS implementation assessments. Results of this study repeat the observation. In addition, results of this study suggest that the most appropriate CAMM archetype is 'Adoption with Benefits'.

Even rigorously utilizing log-based register data of the Kanta Services, however, must accompany parallel and independently conducted HIS/HIE research (e.g. general public and/or professionals' surveys) in order to gain insights and experiences of the whole [28]. Together, 'log-based data' and the 'research-based in-depth data' arms likely complement each other, and form a rich database for further elaboration of outcomes, benefits and harms, and other effects to inform development of new functionalities and better user interfaces for both citizen and professional users.

References

[1] Ludwig DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009;78:22–31..
[2] Coiera E. Building a national health IT system from the Middle Out. JAMIA 2009;16:271–273.
[3] Deutsch E, Duftschmid G, Dorda W. Critical areas of national electronic health record programs: is our focus correct? Int J Med Inform 2010;79:211–222.
[4] Flyvbjerg B. What you should know about mega-projects and why: an overview. Project Manage J 2014;45:6–19.
[5] Ross J, Stevenson F, Lau F, et al. Factors that influence the implementation of e-health: a systematic review of reviews (an update). Implement Sci 2016;11:146.
[6] Fennelly O, Cunningham C, Grogan L, et al. Successfully implementing a national electronic health record: a rapid umbrella review. Int J Med Inform 2020;144:104281.
[7] Ellingsen G, Hertzum M, Melby L. The tension between national and local concerns in preparing for large-scale generic systems in healthcare. Computer Supported Cooperative Work 2022; published online.
[8] Sligo J, Gould R, Roberts V, et al. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017;97:86–97.
[9] Lippeweld T, Sauerborn R, Bodart C (Eds.). Design and implementation of health information systems. Geneva: World Health Organization; 2000.
[10] Finn N. Health information exchange: a stepping stone toward continuity of care and participatory medicine. J Particip Med 2011;3:e47.
[11] Bowden T, Coiera E. Comparing New Zealand's 'middle out' health information technology strategy with other OECD nations. Int J Med Inform 2013;82:e87–e95.
[12] Adler-Milstein J, Ronchi E, Cohen GR, et al. Benchmarking health IT among OECD countries: better data for better policy. JAMIA 2014;21:111–116.
[13] Esmaeilzadeh P, Sambasivan M. Health information exchange (HIE): a literature review, assimilation pattern and a proposed classification for a new policy approach. J Biomed Inform 2016;64:74–86.
[14] Robertson A, Cresswel K, Takian A, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. BMJ 2010;341:c4564.
[15] Klecun E, Zhou Y, Kankanhalli A, et al. The dynamics of institutional pressures and stakeholder behavior in national electronic health record implementations: a tale of two countries. J Inform Technol 2019;34:292–332.
[16] Luz R, Carneiro Mussi C, Dutra A, et al. Implementation of large-scale health information systems. Revistra de Gestao 2021;28:106–132.
[17] Rigby M. Evaluation: 16 powerful reasons why not to do it and 6 over-riding imperatives. MEDINFO 2001. IOT Press 2001;1198–1202.
[18] Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. Finnish J eHealth eWelfare 2018;10:381–395.
[19] Jormanainen V, Parhiala K, Niemi A, et al. Half of the Finnish population accessed their own data: comprehensive access to personal information online is a corner-stone of digital revolution in Finnish health and social care. Finnish J eHealth eWelfare 2019;11:298–310.
[20] Kanta Services. Web pages in English. Accessed January 5, 2022.
[21] Price M, Lau F. The clinical adoption meta-model: a temporal meta-model describing the clinical adoption of health information systems. BMC Med Inform Decision Mak 2014;14:43.
[22] Price M. Clinical Adoption Meta-Model. In: Lau F, Kuziemsky C (Eds). Handbook of eHealth evaluation: an evidence-based approach. Victoria (B.C., Canada): University of Victoria 2016;77–91.
[23] Greenhalgh T, Bowden T. Moving beyond the cathedral model of IT development. BMJ 2011;343:d6480.
[24] Morrison Z, Robertson A, Cresswell K, et al. Understanding contrasting approaches to nationwide implementation of electronic health record systems: England, the USA and Australia. J Healthcare Engin 2011;2:25–41.
[25] Eason K, Dent M, Waterson P, et al. Bottom-up and middle-out approaches to electronic patient information systems: a focus on healthcare pathways. Inform Prim Care 2012;20:51–56.
[26] Tsai CH, Koch S. Towards a framework for national eHealth evaluation and monitoring: a combined top-down and bottom-up approach using Sweden as example. Stud Health Technol Inform 2019;264:954–958.
[27] Jormanainen V, Reponen J. CAF and CAMM analyses on the first 10 years of national Kanta services in Finland. Finnish J eHealth eWelfare 2020;12:302–315.
[28] Vehko T, Ruotsalainen S, Hyppönen H (Eds.). E-health and e-welfare of Finland: check point 2018. THL Report 7/2019. Helsinki: National Institute for Health and Welfare (THL); 2018.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230109

Optimization of Pre-Ictal Interval Time Period for Epileptic Seizure Prediction Using Temporal and Frequency Features

Abdul Aleem SHAIK GADDA [a,1], Dhanvi VEDANTHAM [a], John THOMAS [b], Yuvaraj RAJAMANICKAM [c], Ramshekhar N MENON [d], Jac Fredo AGASTINOSE RONICKOM [a]

[a] *School of Biomedical Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India.*

[b] *Montreal Neurological Institute, McGill University, Montreal, Canada.*

[c] *Science of Learning in Education Centre (SoLEC), National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore.*

[d] *Department of Neurology, R Madhavan Nayar Center for Comprehensive Epilepsy Care, Cognition Behavioral Neurology, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, Kerala, India.*

ORCiD ID: Abdul Aleem Shaik Gadda https://orcid.org/0000-0002-4917-3303

Abstract. Epilepsy is a neurological disorder characterized by recurrent seizures. Automated prediction of epileptic seizures is essential in monitoring the health of an epileptic individual to avoid cognitive problems, accidental injuries, and even fatality. In this study, scalp electroencephalogram (EEG) recordings of epileptic individuals were used to predict seizures using a configurable Extreme Gradient Boosting (XGBoost) machine learning algorithm. Initially, the EEG data was preprocessed using a standard pipeline. We investigated 36 minutes before the onset of the seizure to classify between the pre-ictal and inter-ictal states. Further, temporal and frequency domain features were extracted from the different intervals of the pre-ictal and inter-ictal periods. Then, the XGBoost classification model was utilized to optimize the best interval for the pre-ictal state to predict the seizure by applying Leave one patient out cross-validation. Our results suggest that the proposed model could predict seizures 10.17 minutes before the onset. The highest classification accuracy achieved was 83.33 %. Thus, the suggested framework can be optimized further to select the best features and prediction interval for more accurate seizure forecasting.

Keywords. Epilepsy, Seizure Prediction, Electroencephalogram, Machine Learning, Automated Detection.

1. Introduction

Epilepsy is a neurological disorder that causes sudden, recurring electrical disruptions in the brain. It affects approximately 50 million people of all ages [1], which makes it the second most common neurological disease. Epilepsy may be caused due to head

[1] Corresponding Author: Abdul Aleem Shaik Gadda, E-mail: shaikgadda.aaleem.bme20@itbhu.ac.in..

trauma, stroke, brain tumors, infections, neurodevelopmental disorders like autism, and genetic factors. Epilepsy leads to episodes of seizures, brief changes in behavior, sensations, and sometimes loss of consciousness. Seizures can also cause changes in behavior, mood, and emotions. Seizure prediction is the process of predicting epileptic seizures and providing advance warning of an impending seizure so that clinicians can take preventative measures [2] to reduce the severity of the seizure or avoid it altogether. It also helps improve seizure control and quality of life, decreases the risk of injury, improves overall health, and improves daily functioning. Seizures can be predicted by detecting the beginning of the pre-ictal state, which denotes the onset of the seizure [3]. Electroencephalogram (EEG), which records abnormal electrical activity in the brain, can capture the different stages of seizure inter-ictal, pre-ictal, ictal, and post-ictal. However, seizure prediction has challenges such as variability of patient data, high false-positive rate, computing resources, limited data availability, and imbalanced classes of pre-ictal and inter-ictal states. Optimizing the pre-ictal interval will better classify the pre-ictal state and provide the prediction horizon for forecasting a seizure as different EEG features perform better at different lengths of interval [4]. In this study, we propose a machine-learning model for predicting epileptic seizures by optimizing the pre-ictal interval time. We have used Extreme Gradient Boosting (XGBoost) machine learning algorithm to classify pre-ictal and interictal states as it has shown better performance for diagnosis of epilepsy in the literature [4,5].

2. Methods

The process pipeline adapted in this study is shown in Figure 1.

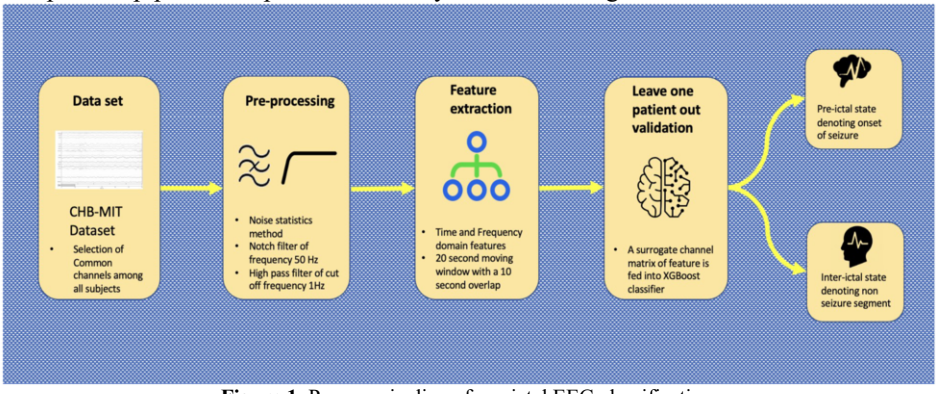


Figure 1. Process pipeline of pre-ictal EEG classification.

In this study, we utilized the CHB-MIT scalp EEG dataset, which is publicly avail-able online. It consists of scalp EEG data of 24 epileptic patients recorded for multiple hours using the conventional 10-20 electrode placement system and split into one hour recordings each. Each file that contained seizure data was annotated at the beginning and end of the seizure. The data was sampled at a frequency of 256 Hz. We found that the number of channels was different between the subjects and varied from 23-40 channels. The 14 channels common among all the subjects were chosen for further analysis. We pre-processed the EEG using a notch filter at 50 Hz and a high-pass filter at 1 Hz to improve the signal-to-noise ratio. The artifact removal was performed using the noise statistics method [5], in which the signal segments with higher RMS value

than a threshold (mean±5*standard deviation) were first eliminated. We extracted an interval of 36 minutes before the onset of the seizure to classify between the pre-ictal and inter-ictal states (figure 2). A moving 20-second window with an overlapping of 10 seconds was chosen for feature extraction. Seizures with a minimum duration of 20 seconds were selected for the study in correspondence to the moving window duration.

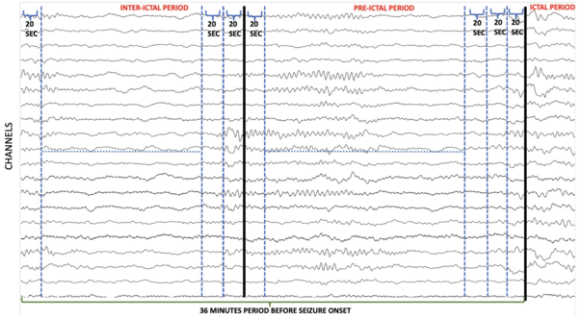


Figure 2. Pre-ictal period optimization process.

We excluded the EEG data of 2 out of 24 subjects as it failed to fit the seizure selection criteria. Time and frequency domain features extracted from the signals are provided in Table 1 [6]. The extracted features were then surrogated into a single channel by averaging the features obtained over all the channels to produce a single-channel feature matrix. These feature matrices obtained were then fed as inputs to the XGBoost classifier [5] as a leave one patient out process in which the classifier was trained, with all the remaining subjects leaving one out for testing. This was repeated for all the possible equal time intervals for the pre-ictal and inter-ictal within 36 minutes of the period before the seizure onset. The final accuracy for an interval was chosen by averaging all the validation accuracies obtained in the leave one patient out method for that particular interval.

Table 1. List of features

Feature type	Features
Time Domain (22)	Absolute Energy, Approximate Entropy, Average Power, Coefficient of Variance, Detrended Fluctuation Analysis (DFA), Higuchi Fractal Dimension, Hjorth Activity, Hjorth Complexity, Hjorth Mobility, Kurtosis, Lempelziv Complexity, Mean, Median, Peak to Peak, Root Mean Square (RMS), Sample Entropy, Skewness, Standard Deviation, Singular Value Decomposition (SVD), Total Signal Range, Variance, Zero Crossing Rate.
Frequency Band Power (3)	Delta Band Power, Beta Band Power, Gamma Band Power.

3. Results

Figure 3 shows the classification accuracy of the XGBoost classifier for the different pre-ictal period intervals. It can be noted that the classification accuracy is minimum at the lower and higher intervals. The classification performance was higher between 6-12 minutes compared to other time periods. We achieved a maximum classification

accuracy of 83.33 % for an epileptic patient. However, the average leave one patient out classification accuracy, recall score, specificity, and F-measure were 61.22 %, 66.92 %, 63.15 %, and 58.85 %, respectively, for a pre-ictal interval having a duration of 10.17 minutes before the seizure onset.

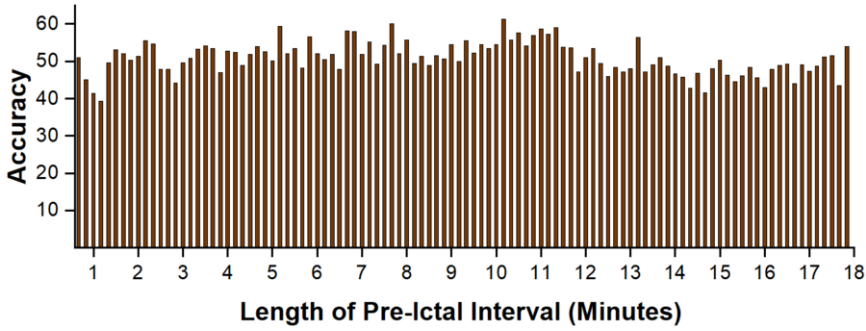


Figure 3. Classification accuracy for different intervals of pre-ictal period.

Figure 4 shows the top 10 features involved in classifying inter-ictal and pre-ictal EEG signals for the interval of 10.17 minutes. We can observe from the figure that the peak to peak and Hjorth complexity were the best features for classifying the pre-ictal and inter-ictal states using EEG signals. This is followed by sample entropy and theta band having influential contributions to classification. Out of the top 10 features, 9 were time domain features.

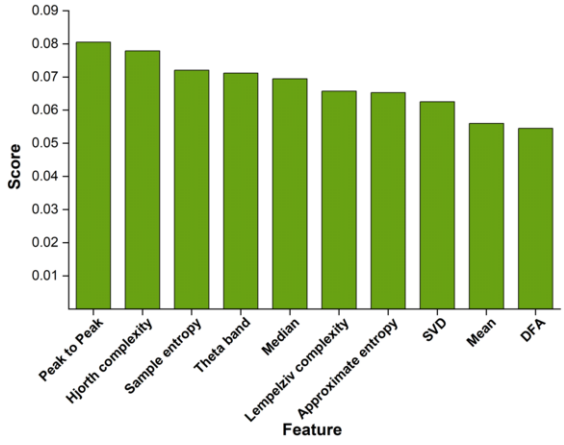


Figure 4. Feature importance based on XGBoost classification.

4. Discussions

We achieved reasonable results in the seizure classification using the CHB-MIT. However, we need to explore our classification model with more data available in other data sets to improve the performance and validate the model's generalizability. Only the time and frequency features were considered for pre-ictal classification and seizure

prediction. We have never employed time-frequency domain features. However, they might provide complex information that might help with better classification. We have only employed the XGBoost model to improve classification performance, but more sophisticated machine learning and deep learning models can be explored. We plan to optimize the model to be able to forecast seizures in real time. In the future, we can expand the research to examine the effects of seizure forecasting on various physiological signals, like ECG and GSR, to study the utility of other biopotential signals on seizure prediction.

5. Conclusions

In this study, a process pipeline was proposed for classifying pre-ictal and inter-ictal states and identifying the pre-ictal period to mark the onset of the seizure. We used the time and frequency domain features to extract the patterns of seizure before onset and XGBoost to build the classification model. The highest classification accuracy achieved was 83.33 % for a pre-ictal duration of 10.17 minutes before the seizure onset. This model in future, has the ability to be deployed as a wearable device for continuous monitoring of epileptic patients.

Acknowledgements

The authors would like to thank the support of the PARAM Shivay supercomputer facility of IIT(BHU), Varanasi, India, used for this research.

References

[1] Fisher RS, Boas WV, Blume W, Elger C, Genton P, Lee P, Engel Jr J. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia. 2005 Apr;46(4):470-2.

[2] Guenot M. Surgical treatment of epilepsy: outcome of various surgical procedures in adults and children. Revue neurologique. 2004 Jun 1;160:5S241-0.

[3] Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, D'Souza W, Yerra R, Archer J, Litewka L, Hosking S. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. The Lancet Neurology. 2013 Jun 1;12(6):563-71.

[4] Thangavel P, Thomas J, Sinha N, Peh WY, Yuvaraj R, Cash SS, Chaudari R, Karia S, Jing J, Rathakrishnan R, Saini V. Improving automated diagnosis of epilepsy from EEGs beyond IEDs. Journal of Neural Engineering. 2022 Nov 24;19(6):066017.

[5] Abirami S, Thomas J, Yuvaraj R, Agastinose Ronickom JF. A Comparative Study on EEG Features for Neonatal Seizure Detection. InBiomedical Signals Based Computer-Aided Diagnosis for Neurological Disorders 2022 (pp. 43-64). Springer, Cham.

[6] Vidaurre C, Krämer N, Blankertz B, Schlögl A. Time domain parameters as a feature for EEG-based brain-computer interfaces. Neural Networks. 2009 Nov 1;22(9):1313-9.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230110

DeepTSE: A Time-Sensitive Deep Embedding of ICU Data for Patient Modeling and Missing Data Imputation

Michael FUJARSKI[a,1,2], Christian PORSCHEN [b,2], Lucas PLAGWITZ[a], Daniel STROTH[a], Catharina Marie VAN ALEN[a], Mahan SADJADI[b], Raphael WEISS[b], Alexander Zarbock[b], Thilo VON GROOTE[b,3] and Julian VARGHESE[a,3]
[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Münster, Münster, Germany*

Abstract. Missing data is a common problem in the intensive care unit as a variety of factors contribute to incomplete data collection in this clinical setting. This missing data has a significant impact on the accuracy and validity of statistical analyses and prognostic models. Several imputation methods can be used to estimate the missing values based on the available data. Although simple imputations with mean or median generate reasonable results in terms of mean absolute error, they do not account for the currentness of the data. Furthermore, heterogeneous time span of data records adds to this complexity, especially in high-frequency intensive care unit datasets. Therefore, we present DeepTSE, a deep model that is able to cope with both, missing data and heterogeneous time spans. We achieved promising results on the MIMIC-IV dataset that can compete with and even outperform established imputation methods.

Keywords. MIMIC IV, ICU, Machine Learning, Deep Embedding, Time-Sensitive Data Imputation, Patient Modeling

1. Introduction

In the intensive care unit (ICU), patient data, such as vital signs or lab values, are often collected at irregular time intervals, resulting in sparse data with many missing values. This is further complicated by time intervals in which data collection is not possible due to the patient not being present in the ICU, due to interventional or diagnostic procedures, or because measuring devices (such as invasive monitoring devices or catheters) are not yet placed. This sparsity of data can make it challenging to accurately assess patient status and make informed treatment decisions. Finally, incomplete datasets prohibit the use of machine learning methods that assume a dataset with no missing values. In addition, traditional

[1] Corresponding Author: Michael Fujarski, E-mail: michael.fujarski@uni-muenster.de
[2] the authors contributed equally to this work and share the first authorship
[3] the authors contributed equally to this work and share the coordinating authorship

statistical methods may not be well-suited to analyze sparse data, as they often rely on the assumption of a sufficient number of observations. To address this data sparsity in the ICU, advanced data imputation techniques or machine learning methods may be necessary to effectively utilize the available data and make accurate predictions. While some methods have been developed in this regards, many of them have significant limitations, such as relying on randomness of missing values, correlations between features, or equidistant time spans. As shown by Sharafoddini et al. missing values do not necessarily occur at random and can even have a predictive potential for patients' 30-day mortality [1].

Time-sensitive deep embeddings and auto-encoders can directly learn from missing values and may therefore be well-suited methods to impute data. They address the issue of data sparsity, as well as feature correlation and relation of features that are not missing by chance. Therefore, we investigated whether and how a variational autoencoder with time-sensitive imputation competes with established methods of data imputation and may therefore yield opportunity for use in clinical data science in the ICU setting.

2. Methods

2.1. Dataset

We used the Medical Information Mart for Intensive Care (MIMIC)-IV database for training and evaluation of the different imputation methods and models2]. MIMIC is a large database of electronic health records (EHRs) of ICU patients at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. It contains a wide range of data including demographics, laboratory test results, medication administrations, and vital sign measurements for more than 40,000 patients. The parameters we considered for this analysis are split into three categories:

The demographics features consist of age, weight, height, and gender of patients at admission. These features are considered as constant during the stay in the ICU.

The vital signs consist of the parameters temperature, respiratory rate, heart rate, mean blood pressure (MBP), urine output as well as the laboratory parameters creatinine, glucose, hemoglobin (Hb), lactate, pH of blood, potassium, oxygen saturation (SpO2), sodium, and the white blood cell count (WBC).

The events are currently limited to the administration of epinephrine, fentanyl, metronidazole, midazolam, morphine, norepinephrine, paracetamol, and propofol. The procedures currently focus on the artificial ventilation of the patients and include the positive end-expiratory pressure (PEEP), the plateau pressure, and the tidal volume.

All EHRs were split into a training (27,000), validation (7,000), and test (8700) set. For every patient, a single time point was defined based on the duration of the ICU stay. On average, the used EHRs were at 71 hours into intensive care for the training set (validation: 71.25, test: 71.5) with a standard deviation of 91.5 (96.5 and 98.25 respectively). The data were used up to the defined point in time per patient. Missing values were replaced by a token depending on the used imputer.

2.2. Imputation

The performance of an imputation method is measured on the test set by masking each parameter separately and calculating the mean absolute error of the available data.
We compare a set of imputation methods with different underlying assumptions and complexity. Simple imputation methods with mean and median assume the data to be normally distributed and to be missing completely at random. The replacement value can be calculated once on a training set resulting in a fast imputation method. Mean and median imputations were done using pandas (1.5.1) directly. It solely depends on populations distribution of the missing value and does not consider any other parameters. More sophisticated approaches like the k-nearest neighbors (KNN) imputer consider all available data to impute values based on a neighborhood. Since KNN strongly depends on the absolute distances, we scaled the data using a standard scaler. Both the standard scaler, and the KNN imputer, were used from scikit-learn (1.1.1). Therefore, it assumes that similar patients, measured as a distance on the available data, can be used to calculate a (weighted) mean of the found neighborhood. Nevertheless, calculating the nearest neighbors on a large dataset can be time-consuming. The multiple imputation by chained equation (MICE) calculates various imputations by modeling the features as functions of other features in a round robin fashion [3]. Further, we propose a Deep Model as another possible imputer.

2.3. Deep Variational Autoencoder

We developed a Variational Autoencoder (VAE) for the ICU simliar to Pinheiro Cinelli et al. [4]. VAEs encode statistical variance into the latent space by modeling learned dimensions as probability functions. However, we expanded the architecture, as shown in Figure 1, to directly learn from missing values as well as encode heterogeneous time spans. Demographic features are encoded as two-dimensional vectors with their actual value and a flag indicating a masked value. Unavailable or masked data are replaced by [0, 1] per feature per patient. Other parameters are three-dimensional vectors, consisting of their value, a timestamp in minutes since the measurements, and a mask. Therefore, the model is capable of handling differently spaced time series. Additionally, our model contains a projection from the latent space into the latent space that predicts the vital signs of a patient for a given positive delta t. The model was trained using a reconstruction loss, predictive loss, and the Kullback-Leibler divergence as implemented by torch (1.12.1). In order to estimate and improve the reconstruction loss on the masked data, we artificially masked the input data. For each epoch, 70% of the input samples were randomly chosen and one random demographic or vital feature was masked. The reconstruction loss was calculated on both the available data as well as the artificially masked input. The predictive loss was calculated during each epoch with 14 out of 18 randomly chosen features to avoid overfitting the model. We further introduced a local Kullback-Leibler divergence on the learned distribution of the input data in order to force an overall lower variance on similar samples. The model was trained for a total of 45,000 epochs. Hyperparameter optimization and model selection were performed on a validation set.

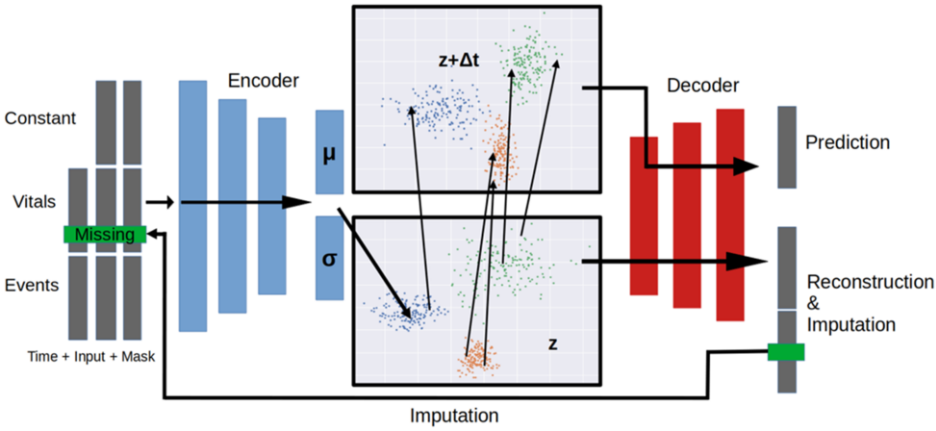


Figure 1. DeepTSE. A Variational Autoencoder with time-sensitive imputation capability. Training losses are calculated on reconstruction, prediction, and local and global Kullback-Leibler divergence.

Performances were evaluated using the separate test set.

3. Results

DeepTSE outperformed other imputation methods in 9 out of 18 categories. Iterative imputation with MICE performed best in 6 categories. Imputation with the train median still performed best in 3 categories, while mean and KNN imputations did not improve the imputation in any case. DeepTSE improved the imputation for the height, weight, gender, pH, creatinine, lactate, urine output, sodium, and WBC. MICE performed best for age, respiratory rate, MBP, heart rate, Hb, and potassium. Median imputation best performed on the body temperature, SpO2, and glucose. Figure 2 depicts the critical differences diagram calculated on the MAE ranking of the different methods. In comparison to the other imputation methods, DeepTSE demonstrated superior performance in 9 out of 18 categories. KNN imputation was outperformed in 14 out of 18 categories, but overall produced only slightly less accurate results to those of MICE. MICE consistently outperformed KNN in the majority of categories and outperformed all other imputation methods in 6 categories. Moreover, MICE is a computationally efficient method, as it is an eager learner, resulting in faster transformations after an initial training. In contrast to the other methods, mean and median imputation methods yielded consistent results across categories, with median imputation performing slightly better than mean imputation.

4. Discussion and Conclusion

DeepTSE has proven to be a well-suited imputer for missing data in the ICU. The combination of time encoding and masking of features enabled the model to learn feature correlations as well as the systematically missing values. Nevertheless, median imputation performed better in three categories indicating that models tend to overfit to features that are missing at random [5]. The imputation of laboratory

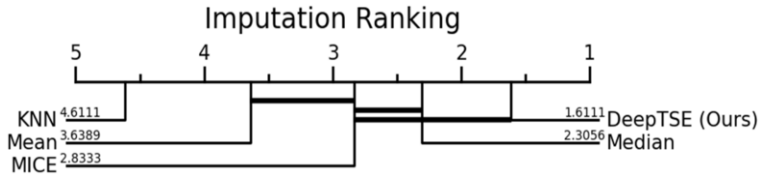


Figure 2. Critical Differences Diagram of the imputation methods and their performances.

values such as lactate and creatinine are particularly of high value for deep learning models, as well as medical professionals. These values are often missing or taken at irregular time points, due to the patient not being present at the ICU or undergoing interventional procedures. Nevertheless, since such values are measured less frequently compared to e.g. vital parameters in an EHR, the loss of information for a potential model is high. If these imputations could be used for estimations of future values, these laboratory measurements could be targets of high interest as well. The early detection of clinical deterioration, as indicated by parameters of lactate or creatinine, are of high importance to clinicians to initiate adequate treatment. For example, serum creatinine is an indicator of acute kidney injury (AKI), the most frequent complication in the critically ill patients, associated with high morbidity and mortality [6]. Similarly, lactate, among other aspects, is an indicator of critical illness as it suggests microvascular dysfunction and lack of peripheral oxygen supply, which is often a sign of hemodynamic or septic shock.

DeepTSE improved and outperformed various imputation methods in 9 out of 18 categories. On average, DeepTSE improved the imputation resulting in a robust and consistent model. Further analyses regarding possible classification and prediction tasks with the learned embedding are now required. Nevertheless, DeepTSE needs to be evaluated on a larger set of parameters, as well as other datasets.

References

[1] Sharafoddini A, Dubin JA, Maslove DM, Lee J. A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study. JMIR Med Info. 2019 Jan.
[2] Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV. Physionet. 2022 Nov.
[3] van Buuren S, Groothuls-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011 Dec.
[4] Cinelli LP, Marins MA, da Silva EAB, Netto SL. Variational Autoenvoder. In Variational Methods for Machine Learning wit hApplications to Deep Networks.; 2021. p. 111-149.
[5] Carpenito T, Manjourides J. MISL: Multiple imputation by super learning. Stat Methods Med Res. 2022 Oct.
[6] Hoste EAJ, Bagshaw SM, Webb S, Kellum JA. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med. 2015 Jul.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230111

Adolescents Identifying Errors and Omissions in Their Electronic Health Records: A National Survey

Josefin HAGSTRÖM[a,1], Charlotte BLEASE[b], Anna KHARKO[b], Isabella SCANDURRA[b] and Maria HÄGGLUND[a]
[a] *Department of Women's and Children's Health, Uppsala University, Sweden*
[b] *Informatics, School of Business Örebro University, Örebro, Sweden*

Abstract. Patient accessible electronic health records (PAEHRs) have been proposed as a means to improve patient safety and documentation quality, as patients become an additional source to detect mistakes in the records. In pediatric care, healthcare professionals (HCP) have noted a benefit of parent proxy users correcting errors in their child's records. However, the potential of adolescents has so far been overlooked, despite reports of reading records to ensure accuracy. The present study examines errors and omissions identified by adolescents, and whether patients reported following up with HCPs. Survey data was collected during three weeks in January and February 2022 via the Swedish national PAEHR. Of 218 adolescent respondents, 60 reported having found an error (27.5%) and 44 (20.2%) had found missing information. Most adolescents did not take any action upon identifying an error or an omission (64.0%). Omissions were more often perceived as serious than errors. These findings call for development of policy and PAEHR design that facilitates reports of errors and omissions for adolescents, which could both improve trust and support the individual's transition into an involved and engaged adult patient.

Keywords. Adolescents; Patient-accessible electronic health records; Errors; Electronic health records; National survey

1. Introduction

To err is human. Patient-accessible electronic health records (PAEHR), which include clinicians' narrative notes, medication lists, and diagnoses, often include errors, possibly leading to safety hazards [1,2]. As EHRs become increasingly used in decision-making and assessment [3], risks include medication errors [4], and delays or missed diagnoses. However, online record access (ORA) can enable patients to be more involved in their care, for example by ensuring accuracy of information [5]. In this paper, the terms error, mistake, and inaccuracy are used interchangeably; similarly, omission and missing information hold the same meaning.

Parent proxy users have been recognized by healthcare professionals (HCPs) as able to detect inaccuracies in their child's EHR [6,7]. In a study that examined error reports among patients and families in a pediatric hospital [8], 35% reported an error. Still, topic

[1] Corresponding Author: Josefin Hagström, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, 751 85 Uppsala, Sweden; E-mail: josefin.hagstrom@kbh.uu.se.

experts have also posited the importance of engaging adolescents in their care to support the transition into adult patienthood [9].

While policies and access ages differ across countries and clinical contexts [10], the transition from parent proxy access into individual access commonly occurs during adolescence. The current variety in implementation largely stems from potential ethical dilemmas, as HCPs worry about maintaining adolescents' confidentiality while providing sensitive care for mental health problems, substance abuse, and sexual health [9]. These uncertainties have led to a lack of studies of how adolescents in fact experience ORA [11]. An interview study with adolescents with cancer and blood disorders who had ORA, identified an appreciation for being able to review updates to the record and ensure its accuracy [12]. However, no studies have yet investigated to what extent specifically adolescents find incorrect and missing information in the PAEHR and what they consequently choose to do. In Sweden, the national PAEHR *Journalen* is available by default to citizens from the age of 16, yet it is possible to apply for access from the age of 13. Our aim is to examine adolescents' experiences of identifying errors and omissions in their records.

2. Methods

To answer the research question, we analyzed a subset of items from the NORDeHEALTH 2022 Patient Survey [13]. The full survey consisted of 83 items exploring the opinions of PAEHR users in Norway, Sweden (ethical approval EPN 2021/05229), Finland and Estonia. This paper only analyzed the data from Sweden. Participants were recruited through the national PAEHR *Journalen*. Upon login, patients received a request for voluntary survey participation together with information about the study. Thus, only active PAEHR users were invited to participate; it was not possible to access the survey without logging in. For this paper, only survey respondents between 15-19 years old were included. Though default access is provided at the age of 16, it is possible to apply for access from the age of 13. However, due to an ethical-legal requirement of written parental consent for research participants younger than 15 years old [14], which would complicate survey distribution, those aged 13 and 14 were excluded prior to data collection. For three weeks from January to February 2022, the survey was available to users accessing their EHR via the Swedish national patient portal. Four closed-ended questions from the survey were included in this study.

1. Have you ever found anything in your record you thought was wrong?
2. Have you ever found anything in your record you thought was missing?
3. If yes, did you do any of the following when you found a mistake or missing information in your record?
4. How easy (or difficult) is it for you to notice errors/mistakes in your record?

The third question had an option for a free-text response ('Something else'). The fourth question was assessed on a 1-5 Likert scale from 'very difficult' to 'very easy'. Data was collected on gender (male, female, other) and age. The collected data were managed by Inera AB, the organization managing the national patient portal and PAEHR *Journalen* (using the survey tool Webropol (version 3.0)). The variables were presented descriptively and gender differences were analyzed using a Fischer's exact test.

3. Results

Of 13 008 respondents, 218 were between 15-19 years old (1.7%). Two respondents were excluded due to being younger than 15. Of the 218 adolescents, 60 reported having found an error (27.5%) and 44 (20.2%) reported missing information (Figure 1). There was no statistical difference among genders in finding errors (Fischer's Exact Test, p=.074), yet, 63.6% (7/11) of those that marked gender as "other" reported finding an error (17.9% (7/39) among males and 27.4% (46/168) among females). No gender difference was found for omissions (Fischer's Exact Test, p=.322).

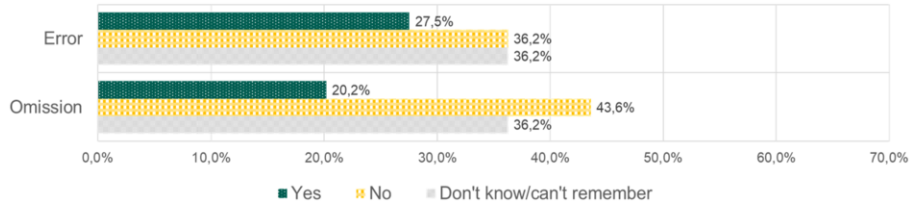


Figure 1. Errors and omissions found by adolescents aged 15-19 (n=218).

Of the adolescents who found an error, one third (20/60) perceived it as somewhat or very serious (Figure 2). While none rated omissions as not at all serious, almost 89% (39/44) assessed them as somewhat or very serious. A larger proportion of respondents were uncertain of the gravity of omissions (5/44, 11.4%) than of errors (2/60, 3.3%).

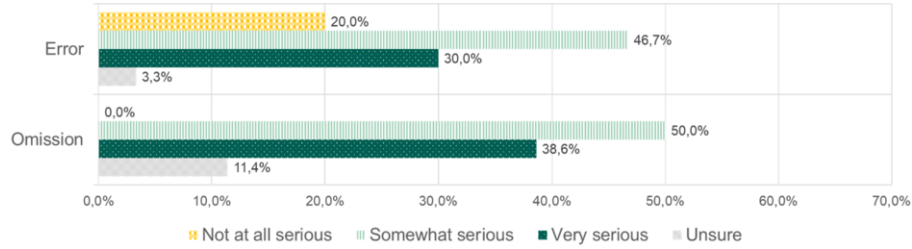


Figure 2. Adolescents' perception of the gravity of found errors (n=60) and omissions (n=44).

Upon finding errors or omissions, adolescents most frequently reported having done nothing (55/86, 64.0%). Informing the HCP at the next visit or contacting the healthcare unit via telephone was reported by 14.0% respectively (12/86) (Figure 3).

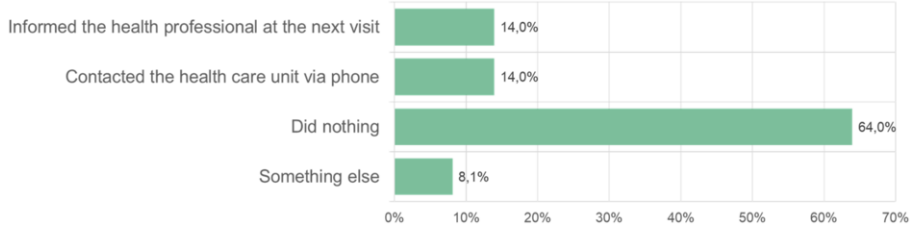


Figure 3. Action taken by adolescents when finding errors or omissions in the PAEHR (n=86).

Seven respondents selected "Something else" (8.1%) and described asking a nurse, informing other HCPs, leaving a complaint, plans to inform future HCPs, plans on applying for the full physical copy of the record as well as contacting the "contact point". Respondents rated the ease of finding errors or omissions at a mean of 2.64 (SD=1.25).

4. Discussion

Our survey collating 218 adolescent respondents is, to our knowledge, the first systematic examination of adolescents' identifying mistakes in their PAEHR. One fourth of adolescents had identified an error, and one fifth had found omissions. The omissions were more often rated as serious than errors were. A larger proportion of those marking their gender as "Other" reported finding errors in their record than among males and females. This is in line with a recent study where adult transgender people reported recurrently finding errors in their records, such as the wrong name or pronoun [19].

Previous research has suggested that patients hold interest in engaging in safety efforts [15]. Notwithstanding, it was found that compared to the 28% who reported having informed HCPs via phone or during a visit, a majority (64%) of adolescents did nothing when finding a mistake. This may be attributed to the fact there is currently no guidance or instruction in *Journalen* on how to report errors or omissions. Only upon searching the matter online, the general advice is to contact the healthcare provider, which in itself may be a daunting process. Furthermore, since research has found that anonymous reporting systems appear less challenging and confrontational for patients in general [16], it is possible that adolescent patients hesitate to notify HCPs of errors or omissions directly. In addition, respondents may fear repercussions or prefer to report adverse events via electronic media, as reported by a focus group study with adolescent patients based in the United States [17]. Respondents rated the ease of detecting errors and omissions as lower than average. A possible explanation for this is that adolescents appear to have lower self-esteem than other age groups [18], they feel insecurity in clinical environments and may lack health literacy compared with adults [17].

The study has limitations. The sample size was small, which reflects difficulties in recruiting adolescents in research. Data on socioeconomic background or healthcare needs was not collected, limiting our analysis. Furthermore, the terminology may have been confusing, as missing information can be perceived as a type of error. This may have affected participants' responses. The quantitative analysis presented in this study does not give us insight into what type of errors or omissions adolescents find. Also, the clinical relevance of the errors and omissions identified by adolescents cannot be determined. Future studies should further explore these issues. Work is also encouraged in other settings, as patients' expectations on EHRs may differ according to culture [20] and healthcare system. For example, EHR documentation and patient perception of errors are likely to be different in the United States compared to Sweden, due to a larger focus on issues such as insurance coverage and risk of litigation.

5. Conclusions

With the implementation of patient access to their records, accuracy of EHRs is not only critical for maintaining quality of care, but to maintain patients' perception of quality and their trust in healthcare. This study indicates that adolescents can play an important part in identifying errors and omissions in their records, but improved processes for reporting errors and omissions are needed to facilitate adolescents' contribution to improved patient safety. The findings call for development of policy and PAEHR design that facilitates reports of errors and omissions for adolescents, which could improve trust and support the individual's transition into an involved and engaged adult patient.

6. Acknowledgement

This work was supported by NordForsk through funding to Nordic eHealth for Patients: Benchmarking and Developing for the Future (NORDeHEALTH), number 100477.

References

[1] Ely JW, Kaldjian LC, D'Alessandro DM. Diagnostic Errors in Primary Care: Lessons Learned. J Am Board Fam Med 2012;25:87–97, doi: 10.3122/jabfm.2012.01.110174.
[2] Singh H, Giardina TD, Meyer AND, Forjuoh SN, Reis MD, Thomas EJ. Types and Origins of Diagnostic Errors in Primary Care Settings. JAMA Intern Med 2013;173:418, doi:10.1001/jamainternmed.2013.2777.
[3] Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using Clinical Notes and Natural Language Processing for Automated HIV Risk Assessment. J Acquir Immune Defic Syndr 1999 2018;77:160–6, doi: 10.1097/QAI.0000000000001580.
[4] McDonald CJ. Computerization Can Create Safety Hazards: A Bar-Coding Near Miss. Ann Intern Med 2006;144:510, doi: 10.7326/0003-4819-144-7-200604040-00010.
[5] Blease CR, Bell SK. Patients as diagnostic collaborators: sharing visit notes to promote accuracy and safety. Diagnosis 2019;6:213–21, doi: 10.1515/dx-2018-0106.
[6] Chung RK, Kim UO, Basir MA. Differing perspectives on parent access to their child's electronic medical record during neonatal intensive care hospitalization: a pilot study. J Matern Fetal Neonatal Med 2018;31:1078–84, doi: 10.1080/14767058.2017.1306853.
[7] Kelly MM, Coller RJ, Hoonakker PLT, Nacht CL, Dean SM. Provider Experiences With Offering Families Bedside Health Record Access Across a Children's Hospital. Hosp Pediatr 2020;10:1002–5, doi: 10.1542/hpeds.2020-0044.
[8] Bourgeois FC, Fossa A, Gerard M, Davis ME, Taylor YJ, Connor CD, et al. A patient and family reporting system for perceived ambulatory note mistakes: experience at 3 U.S. healthcare centers. J Am Med Inform Assoc 2019;26:1566–73, doi: 10.1093/jamia/ocz142.
[9] Bourgeois FC, DesRoches CM, Bell SK. Ethical Challenges Raised by OpenNotes for Pediatric and Adolescent Patients. Pediatrics 2018;141:e20172745, doi: 10.1542/peds.2017-2745.
[10] Hagström J, Scandurra I, Moll J, Blease C, Haage B, Hörhammer I, et al. Minor and Parental Access to Electronic Health Records: Differences Across Four Countries. Stud Health Technol Inform 2022:495–9, doi: 10.3233/SHTI220508.
[11] Hagström J, Blease C, Haage B, Scandurra I, Hansson S, Hägglund M. Use, Views of and Experiences of Web-Based Access to Electronic Health Records for Children, Adolescents, and Parents: Scoping Review. J Med Internet Res 2022, doi: 10.2196/40328.
[12] Hong MK, Wilcox L, Feustel C, Wasileski-Masker K, Olson TA, Simoneaux SF. Adolescent and Caregiver use of a Tethered Personal Health Record System. AMIA Annu Symp Proc AMIA Symp 2016;2016:628–37.
[13] Hägglund M. Nordic countries lead new initiative on patient access to EHRs. TheBMJopinion 2021. https://blogs.bmj.com/bmj/2021/05/18/maria-hagglund-nordic-countries-lead-new-initiative-on-patient-access-to-ehrs/ (accessed January 16, 2022).
[14] Riksdagsförvaltningen. 2003:460. 2003.
[15] Mohsin-Shaikh S, Garfield S, Franklin BD. Patient involvement in medication safety in hospital: an exploratory study. Int J Clin Pharm 2014;36:657–66, doi: 10.1007/s11096-014-9951-8.
[16] Davis RE, Sevdalis N, Vincent CA. Patient involvement in patient safety: How willing are patients to participate? BMJ Qual Saf 2011;20:108–14, doi: 10.1136/bmjqs.2010.041871.
[17] Sawhney PN, Davis LS, Daraiseh NM, Belle L, Walsh KE. Barriers and Facilitators of Adverse Event Reporting by Adolescent Patients and Their Families. J Patient Saf 2020;16:232–7, doi: 10.1097/PTS.0000000000000290.
[18] Orth U, Robins RW. The Development of Self-Esteem. Curr Dir Psychol Sci 2014;23:381–7, doi: 10.1177/0963721414547414.
[19] Alpert AB, Mehringer JE, Orta SJ, Redwood E, Hernandez T, Rivers L, et al. Experiences of Transgender People Reviewing Their Electronic Health Records, a Qualitative Study. J Gen Intern Med 2022, doi: 10.1007/s11606-022-07671-6.
[20] van den Brink-Muinen A, Verhaak P, Bensing J, Bahrs O, Deveugele M, Gask L, et al. Communication in general practice: differences between European countries. Fam Pract 2003;20:478–85, doi: 10.1093/fampra/cmg426.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230112

The Smart Data Extractor, a Clinician Friendly Solution to Accelerate and Improve the Data Collection During Clinical Trials

Sophie QUENNELLE[a,b,c,1], Maxime DOUILLET [d], Lisa FRIEDLANDER[b,c], Olivia BOYER [b,c], Antoine NEURAZ[a,b,c], Anita BURGUN[a,b,c] and Nicolas GARCELON[a,b,d]

[a] *HeKA Team, Inria Inserm UMR_S1138, PariSantéCampus, Paris, France*
[b] *Université de Paris Cité, Paris, France*
[c] *Hôpital Universitaire Necker-Enfants malades, APHP, Paris, France*
[d] *Data Science Platform, Imagine Institute, Paris, France*

Abstract. In medical research, the traditional way to collect data, *i.e.* browsing patient files, has been proven to induce bias, errors, human labor and costs. We propose a semi-automated system able to extract every type of data, including notes. The Smart Data Extractor pre-populates clinic research forms by following rules. We performed a cross-testing experiment to compare semi-automated to manual data collection. 20 target items had to be collected for 79 patients. The average time to complete one form was 6'81'' for manual data collection and 3'22'' with the Smart Data Extractor. There were also more mistakes during manual data collection (163 for the whole cohort) than with the Smart Data Extractor (46 for the whole cohort). We present an easy to use, understandable and agile solution to fill out clinical research forms. It reduces human effort and provides higher quality data, avoiding data re-entry and fatigue induced errors.

Keywords. Electronic Health Records, Clinical Research Forms, Clinical Data Reuse, Observational Study

1. Introduction

Most of the patients' information required for clinical trials and registries are available in patients' electronic health records (EHRs).[1] The most common manner to fill Case Report Forms (CRF) is still to browse patients' documents searching for the information required by the study protocol. This process induces delays, human efforts, costs, and risks of transcription errors. Recent efforts have been dedicated to reuse EHR data to identify patients eligible for trials to optimize clinical trial protocols and to transcribe the variables of interest from EHRs to CRFs automatically.[2, 3] However, several pitfalls remain since EHR data are heterogeneous, completeness of structured data elements is low and most of the clinical information is locked into medical notes and needs to be transformed in a structured format before secondary use.[4] Our objective was to develop

[1] Corresponding Author : Sophie Quennelle, E-mail: sophie.quennelle@protonmail.com

a pipeline able to speed up the collection of data required by a CRF, from all document sources, and to support user-friendly data quality assessment. We evaluated this tool through a retrospective study.

2. Method

2.1. Material

Necker-Enfants Malades Hospital is an AP-HP university children's hospital in Paris with a data analytics and warehousing solution, called Dr. Warehouse (DRWH) [5]. It integrates multiple data sources ranging from structured data to free text clinical narratives and applies natural language processing methods to medical text to detect negation, family history, and to extract phenotypic information based on the Unified Medical Language System Metathesaurus®.

2.2. Methods

The Smart Data Extractor (SDE) has been developed on top of DRWH to help researchers with patient information retrieval and CRF completion. The SDE is adapted as follows to populate a given CRF (Figure 1): (1) The user provides a formal representation of each item of the research protocol (name, type, list of accepted values) and associates specific extraction rules to each item. The items and queries can be imported or created by an expert. If the data of interest is stored in a structured format, the rule includes the corresponding thesaurus codes (e.g. LOINC, ICD10, or local nomenclature).

If the data is to be searched for in the clinical notes, the system can simply find the documents containing the items based on the labels of the variables or items in a list and check the corresponding box. The user can also specify the regular expressions (REGEX) that are used to retrieve information from clinical notes. (2) The software works as an automatic search engine mining the patients reports to look for the sentences containing the items and values.

The SDE has been designed to be used in two ways: semi or fully automated. In the fully automated mode, the SDE extracts automatically the completed CRF for each patient of the study cohort. This approach is appropriate when the data of the questionnaire are unambiguous. In the semi-automated mode, the user needs to validate manually each item of the CRF and can modify the answers. This approach is suitable for ambiguous variables that require human validation. We designed a proper Human Machine Interface for this validation. For each item, the SDE screens the patient EHR and suggests variables matching the query. If an answer is found in a negative syntagm, the interface displays this information to the user. As the variables are presented in their context (sentences) the expert is able to select the most appropriate occurrences. The link between the data extracted and the health reports is maintained, to ensure data assessment at any time.

At any time, experts can edit and improve the list of items and queries of the SDE. When the CRF is completed for the cohort, it can be exported into a format suitable for statistical analysis. Authorized users can export the data and/or transfer it to a Research Electronic Data Capture (REDCap) database.[6]

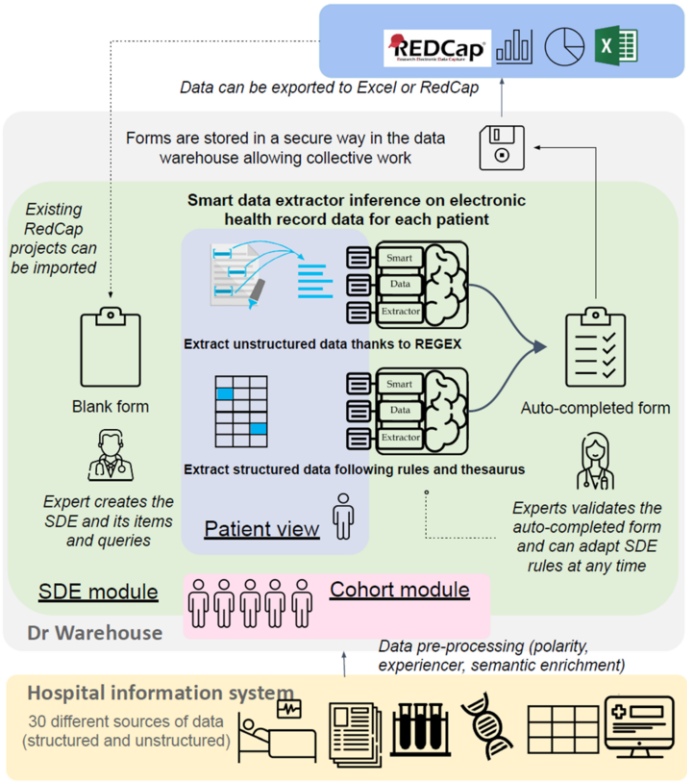


Figure 1. Graphical presentation of the Smart Data Extractor.

2.3. Evaluation

The evaluation protocol consisted in comparing manual versus SDE-assisted data collection of those 20 variables required to calculate a risk score related to the cardiac catheterization procedure. A medical doctor (MED) designed a SDE dedicated to this task after two days of training on REGEX and DRWH Thesaurus. Predefined functions were used for demographic data. Biological and hemodynamics were extracted based on the corresponding thesaurus code. A list of REGEX was defined for unstructured items. When appropriate, time intervals were defined in order to limit the data collection in a specific period around the procedure. We randomly divided 79 patients into two subgroups. Two researchers, a clinical research assistant with expertise in DRWH (TEC) and a cardiologist (MED) conducted the data collection. They filled in the forms in a cross-testing experiment: for the first subgroup, the research assistant completed the CRF manually, i.e., he read every patient's files on the database and filled in the eCRF without any computer assistance while the cardiologist used the SDE. For the second subgroup TEC was assisted by the SDE while MED completed the eCRF manually. The two cohorts were then reconciliated to identify discrepancies, compare the share of missing data and data collection times.

3. Results

Group 1 was composed of 39 patients, with a mean number of documents of 111 and a mean length of follow-up of 2.8 years. Group 2 was composed of 40 patients, with a mean number of documents of 177 and a mean length of follow-up of 3.2 years. The mean data collection time per patient was lower when the users were assisted by the SDE (3'27'' for TEC + SDE on group 2 and 3'17'' for MED + SDE on group 1 versus 6'23'' for TEC alone on group 1 and 7'38'' for MED alone on group 2).

Discrepancies between TEC and MED were comparable for the two groups (103 for group 1 and 106 for group 2). Regardless of the user, there were more mistakes when the form was filled in manually (in total, 163 errors in the manually filled in forms versus 46 in the semi-automated filled in forms). Three types of errors were made: (1) missing values, i.e. data that were not retrieved in patient EHR by one of the researchers, (2) breach of data collection protocol, (3) misunderstanding of an ambiguous type of procedure or of the patient's medical history in text.

We paid attention to the software's display and ergonomics. The patient file and the CRH are displayed side by side allowing the user to read the context of each extracted concept. TEC appreciates the fact that the SDE points out the relevant part of the patient file for each clinical question. MED highlighted the fact that the SDE was able to detect family history and negation which avoids the burden generated through false positives keywords detection.

4. Discussion

We presented and evaluated the SDE, a generic inference engine that automatically populates CRF based on EHRs. Globally, the time needed to fill out the CRF was divided by two and the number of errors by three. The results were similar for the two users (MED or TEC).

Medical data extraction solutions should permit text mining as clinical narratives are the primary source of information about patient history, treatment, and disease course. To avoid the burden of false positive keywords, it is critical to detect negated clinical signs and family medical history. Both automatic and semi-automatic data extraction can lead to errors [7] but the semi-automated method reduces the eventuality that the researchers make two different interpretations of the exact same event, the final human check ensures that the algorithm is working properly and requires fewer human and computer resources to develop [8]. With the SDE the link between each extracted information and its document source and context is maintained, which ensures the possibility of data verification and traceability at any time. A clinical extraction tool must be easily adaptable to the needs of any research protocol while remaining simple enough to be handled by clinicians. Shalhout and al. proposed a pipeline to capture structured clinical data to a REDCap based registry.[9] Miller and al. provide a powerful application to abstract clinico-genomic data from EHR but this extraction solutions are limited to a predefined sets of data.[10] We sought to design the SDE in an intuitive and user centered way so that non informatic trained researchers (nurses, medical students, physicians…) could configure the data collection forms according to their specific needs.

Our evaluation protocol had some limitations: first, it is not possible to discriminate with certainty missing data from data that existed in the EHR but was not identified by either of the two users. Secondly, the time allocated to implement the tool has not been

measured. Thirdly, the rules, conceived by a non-specialist, were probably not optimal. This may have had an impact on the efficiency of our SDE, but it also reflects the real-life use of the tool. Finally, in our specific use case 5 variables only had to be searched in the text which explains the relatively short time needed to complete each patient form. A more tedious CRF would probably have shown a more significant gain with the SDE. To favor the widespread use of the tool, efforts should be made on interoperability. We aim to make the SDE universal and adaptable to any type of hospital database. This could facilitate secure data export and exchange and so multi-center clinical registries.

5. Conclusion

The SDE is an easy-to-use semi-automated data collection system able to extract all variables of interest whatever their format (structured and unstructured) from the patient EHR to fill in CRF. The SDE automatically extracts patients' structured data and assists researchers in text mining for the semi-automatic extraction of data reported by caregivers in clinical notes. We are convinced that the SDE can promote multi-centered trials, reduce costs and clinical research cycle time. The SDE is a semi-automated framework, it requires human effort and validation and does not guarantee zero missing data and error rates but unlike a complex NLP model, it does not necessitate a step of training and can adapt to any clinical subject in a short time.

References

[1] El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic PY, Jaulent MC, Daniel C. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. Journal of biomedical informatics. 2011 Dec 1;44:S94-102. doi:10.1016/j.jbi.2011.07.007.
[2] EHR4CR | Electronic Health Records Systems for Clinical Research, IMI Innovative Medicines Initiative.http://www.imi.europa.eu/projects-results/project-factsheets/ehr4cr
[3] Brat GA, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. NPJ digital medicine. 2020 Aug 19;3(1):109. doi:10.1101/2020.04.13.20059691.
[4] Vass A, Reinecke I, Boeker M, Prokosch HU, Gulden C. Availability of Structured Data Elements in Electronic Health Records for Supporting Patient Recruitment in Clinical Trials. InMEDINFO 2021: One World, One Health—Global Partnership for Digital Innovation: Proceedings of the 18th World Congress on Medical and Health Informatics 2022 Jun 1 (Vol. 290, p. 130). IOS Press. doi:10.3233/SHTI220046.
[5] Garcelon N, Neuraz A, Salomon R, Faour H, Benoit V, Delapalme A, Munnich A, Burgun A, Rance B. A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. J Biomed Inform. 2018 Apr. doi: 10.1016/j.jbi.2018.02.019.
[6] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009 Apr 1;42(2):377-81.. doi:10.1016/j.jbi.2008.08.010.
[7] Byrne MD, Jordan TR, Welle T. Comparison of manual versus automated data collection method for an evidence-based nursing practice study. Applied Clinical Informatics. 2013;4(01):61-74. doi:10.4338/ACI-2012-09-RA-0037.
[8] Cheng AC, Duda SN, Taylor R, Delacqua F, Lewis AA, Bosler T, Johnson KB, Harris PA. REDCap on FHIR: clinical data interoperability services. Journal of biomedical informatics. 2021 Sep 1;121:103871.. doi:10.1016/j.jbi.2021.103871.
[9] Shalhout SZ, Saqlain F, Wright K, Akinyemi O, Miller DM. Generalizable EHR-R-REDCap pipeline for a national multi-institutional rare tumor patient registry. JAMIA open. 2022 Apr;5(1):ooab118. doi:10.1093/jamiaopen/ooab118.
[10] Miller DM, Shalhout SZ. GENETEX—a GENomics Report TEXt mining R package and Shiny application designed to capture real-world clinico-genomic data. JAMIA open. 2021 Jul;4(3):ooab082.. doi:10.1093/jamiaopen/ooab082.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230113

Monitoring Distributed Business Processes in Biomedical Research

Christoph ZILSKE[a], Maximilian KURSCHEIDT[a,1], Simon T SCHWEIZER[a], Hauke HUND[a], Simon MÖDINGER[a] and Christian FEGELER[a]
[a] *GECKO Institute, Heilbronn University of Applied Sciences, Heilbronn, Germany*

Abstract. In biomedical research, business processes, such as data-sharing or feasibility queries, span across several healthcare organizations. Due to the growing number of data-sharing projects and connected organizations, the management of distributed processes gets more complex over time. This leads to an increased need for administrating, orchestrating, and monitoring all distributed processes of a single organization. A proof of concept for a decentralized and use case agnostic monitoring dashboard was developed for the Data Sharing Framework, which most German university hospitals have deployed. The implemented dashboard can handle current, changing, and upcoming processes using only information for cross-organizational communication. This differentiates our approach from other existing use case specific content visualizations. The presented dashboard is a promising solution to provide administrators with an overview of the status of their distributed process instances. Therefore, this concept will be further developed in upcoming releases.

Keywords. Visualization, Decentralized Monitoring, Distributed Business Process, Data Sharing

1. Introduction

Large amounts of medical routine data are documented across different healthcare organizations, which provides great potential for biomedical research. To consolidate and analyze this data, distributed processes such as feasibility studies [1], record linkage, and data sharing [2] need to be implemented between bio-medical research organizations. The requirements to facilitate such distributed processes include process orchestration, technical connectivity, and data protection. In this context, the Data Sharing Framework (DSF) was introduced as a use case independent communication infrastructure for executing cross-organizational distributed processes. Currently, the DSF is deployed at approx. 40 organizations in production, including most German university hospitals and several other organizations across multiple research consortia, such as the *Medical Informatics Initiative* (MII) [3,4] and the *Network University Medicine* (NUM) [5].

The ever-growing number of federated data-sharing projects, respective processes, and participating organizations result in increased communication relationships that require a higher degree of automation. This trend leads to a situation where monitoring the status, tasks, and errors of distributed processes becomes more challenging for users

[1] Corresponding Author: Maximilian Kurscheidt, GECKO Institute, University of Applied Sciences Heilbronn, Max-Planck-Straße 39, 74081 Heilbronn, Germany; E-mail: maximilian.kurscheidt@hs-heilbronn.de.

and administrators. Due to changing requirements and a distributed development community, the exact design of future processes is unknown. With this article, we want to answer the question if a decentralized dashboard is suitable for monitoring the status of distributed processes. Therefore, a proof of concept was implemented.

1.1. Data Sharing Framework

The concept of the DSF represents a peer-to-peer network where no central entity exists, and each organization using the DSF has various communication partners executing distributed processes. The DSF was implemented using commonly known standards such as FHIR R4 and BPMN 2.0. It consists of a FHIR server, accessible by external communication partners, acting as a "mailbox"; and a private Business Process Engine (BPE) executing BPMN 2.0 processes to interact with local and remote systems [3].

The decentralized authentication, authorization, and role definition specify the relationship between organizations and are handled through the DSF-*AllowList*[2], which consists of O*rganization*, *OrganizationAffiliation*, and *Endpoint* FHIR resources.

FHIR *Task* resources are used to start and continue processes and include mandatory attributes in the DSF context: requester; recipient; status of the process; a business key; and the name and version of the process that should be triggered. Input parameters can be used to configure processes and output parameters to save results.

Organizations can read *Task* resources stored on remote FHIR servers if they have created them. They can also access all *Task* resources stored on their local FHIR server. Each *Task* is associated with a specific process instance by their business key, which is shared across all *Task* resources. During the execution of a process, *Task* resources are updated with the status, changing from *requested* to *in-progress* to *completed* or *failed.*

2. Materials and Methods

Requirements for the proof of concept were conducted by four expert interviews, where interviewees had multiple roles (DSF core developers, plugin developers, technical support engineers and system administrators). The interviews included questions regarding the functionality, interactions, challenges and suggested improvements of the DSF and processes. Distributed processes from the research consortia MII and NUM were analyzed regarding communication patterns, entity roles and structure. Additionally, protocols of past DSF hackathons were reviewed for common challenges.

The dashboard was developed for the DSF using common web technologies, such as HTML, CSS, and the JavaScript framework React.js[3]. The DSF was used in version 0.9[2], widely deployed at German university hospital *data integration centers* (DIC).

3. Results

Most organizations communicate simultaneously with other organizations in multiple projects through different processes via the DSF communication infrastructure. An

[2] https://github.com/highmed/highmed-dsf/tree/v0.9.0 (accessed March 3, 2023)

[3] https://reactjs.org (accessed March 3, 2023)

example of such a peer-to-peer network constellation from the perspective of *Organization A* with the role DIC is displayed in *Figure 1*. Each of the edges represents a line of communication associated with a distributed process, distinguished by color.

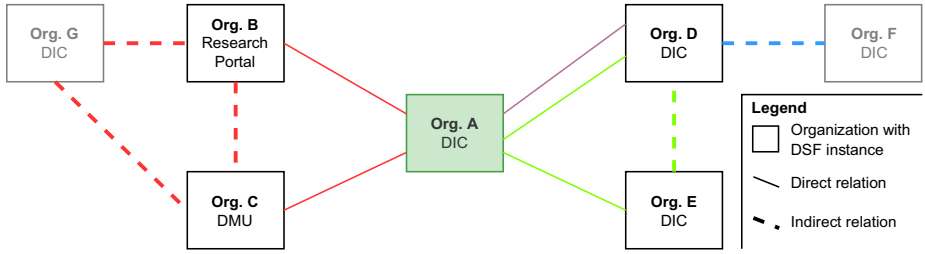


Figure 1. Simplified distributed communication infrastructure with multiple distributed processes. DIC: *Data integration center*; DMU: *Data management Unit*.

The red communication lines in *Figure 1* are derived from the data sharing process of the MII[4] . A simplified version of the BPMN 2.0 process is displayed in *Figure 2* and the model shows the interaction pattern, including multiple subprocesses from different roles. FHIR *Task* resources are represented through the colored BPMN message flows.

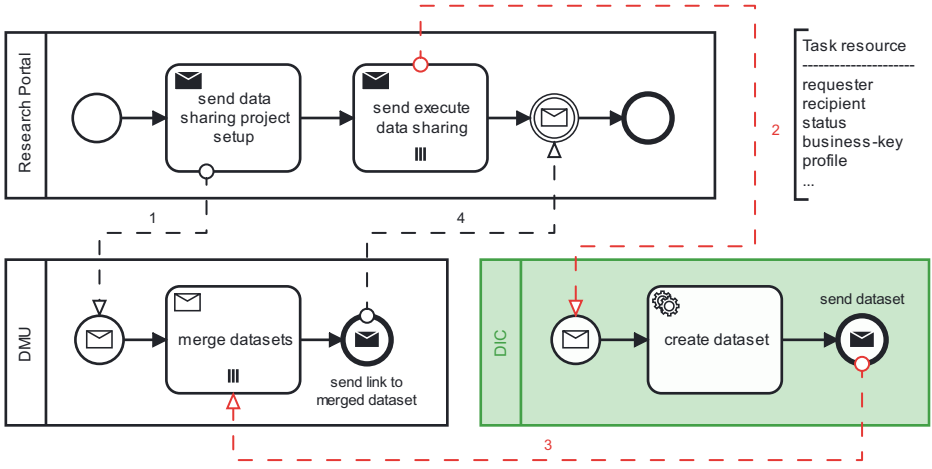


Figure 2. Simplified data sharing process between the roles Research Portal, Data Management Unit (DMU), and Data Integration Center (DIC).

The DIC, in this example, receives a *Task* resource from the *Research Portal* and sends a *Task* resource to the *Data Management Unit* (DMU). Focusing on the perspective of the DIC, only two *Task* resources marked *2* and *3* are accessible for the DIC. *Task* resources *1* and *4* are relevant for the overall process but are executed between other organizations and outside of the DIC's scope. This concept is consistent across all processes developed for the DSF.

The interviewees stated the following key requirements for the dashboard: all deployed processes with their corresponding process instances need to be presented; the

[4] https://github.com/medizininformatik-initiative/mii-dsf-processes/wiki (accessed March 3, 2023)

dashboard should be use case and process agnostic, deployed locally and independent from other participants; received and requested *Task* resources and their associated process instances need to be presented; *Task* resources and process instances should be filterable and searchable.

The information included in the *Task* resources was used to visualize the state of a distributed process in our monitoring dashboard. *Figure 3* shows the implemented monitoring dashboard from the perspective of the DIC referenced in *Figure 1* and *Figure 2*. The elements presented in *Figure 3 (section a)* allow the user to select between different distributed processes deployed on the DSF. If a primary process is selected, all corresponding process instances grouped by their individual business key are listed in *Figure 3 (section b)*.

In *Figure 3*, the simplified data sharing process *from Figure 2* is selected. Selecting a process instance opens an overview of the received and sent *Task* resources marked by the icons in *Figure 3 (section c*). They present their status, which gets aggregated to the overall status of the process instance. *Task* resources and process instances can be sorted and filtered by different properties such as status, requester, or date of creation. This functionality allows a structured overview of the status of process instances. The *Task* resource can be selected to view the entire content.

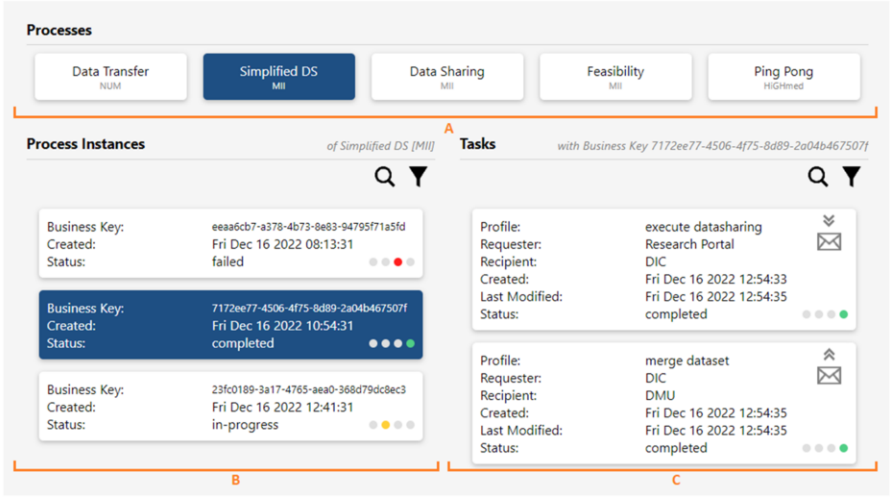


Figure 3. Monitoring dashboard: ***Section a***: Main processes deployed on the DSF. ***Section b***: Process instances of the selected process. ***Section c***: Associated task resources.

4. Discussion

We implemented a decentralized dashboard prototype to monitor the state of distributed processes. It bundles relevant information and visualizes the status of process instances for the system administrator. This eliminates the need to analyze *Task* resources and log files manually. The peer-to-peer architecture of the DSF enables organizations to execute individual processes with their distributed communication partners. No central entity oversees and coordinates the execution of process instances for all organizations. Thus, a decentralized monitoring dashboard was implemented, which enables every organization to operate and monitor their processes independently.

A use case independent approach was chosen to monitor known and unknown processes. This differentiates our approach from centralized use case specific dashboards (e.g., MII *Forschungsdatenportal für Gesundheit* [6]). These are developed to represent detailed process content but are difficult to adapt to other processes. Our approach switches the perspective from use case specific dashboards to an organization-oriented dashboard that visualizes all distributed processes associated with their DSF instance. No adaptations of existing processes or the DSF are needed for our implementation. Moreover, no modification of the specifically secured BPE component is required. Our monitoring approach uses only required *Task* resource attributes, representing communication metadata. An overview of the distributed process was achieved by grouping and labeling the information available within these resources. The level of detail might still be insufficient for users, depending on their role in the process. In future, through a standardized interface, additional information about a process could be integrated into the dashboard.

5. Conclusion

In this article, we presented a proof of concept for a decentralized monitoring dashboard for visualizing distributed processes executed on the communication infrastructure - Data Sharing Framework. Using only information contained in FHIR *Task* resources for cross-organizational communication, we generalized the dashboard functionality to handle existing and future distributed processes. This design decision makes the presented proof-of-concept use case-, process-, project- and consortia independent. The presented decentralized dashboard is a promising solution to provide administrators with an overview of the status of their distributed process instances that are parallelly deployed in their DSF. The dashboard will be further developed as part of the *DSF-Community-Project* during the MII phase 3 grant.

Acknowledgments

The project is funded by the German Federal Ministry of Education and Research (BMBF, grant id: 01ZZ1802E).

References

[1] R. Wettstein et al., Secure Multi-Party Computation Based Distributed Feasibility Queries – A HiGHmed Use Case. Stud Health Technol Inform. 296 (2022) 41-49. doi:10.3233/SHTI220802
[2] R. Wettstein et al. Data Sharing in Distributed Architectures – Concept and Implementation in HiGHmed. Stud Health Technol Inform. 283 (2021) 111-118. doi:10.3233/SHTI210548
[3] H. Hund et al. Executing Distributed Healthcare and Research Processes – The HiGHmed Data Sharing Framework. Stud. Health Technol. Inform. 278 (2021) 126-133. doi:10.3233/SHTI210060
[4] P. Knaup et al., Implementation of a National Framework to Promote Health Data Sharing, Yearb Med Inform, 27.01 (2018) 302-304. doi: 10.1055/s-0038-1641210
[5] H.U. Prokosch et al. The COVID-19 Data Exchange Platform of the German University Medicine. Stud. Health Technol. Inform. 294 (2022) 674–678. doi:10.3233/SHTI220554
[6] J. Gruendner et al. The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. JMIR Med Inform 10:5 (2022). doi: 10.2196/36709

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230114

Low Valence Low Arousal Stimuli: An Effective Candidate for EEG-Based Biometrics Authentication System

Jahanvi JESWANI [a], Praveen Kumar GOVARTHAN [a,1], Abirami SELVARAJ [a], Amalin PRINCE [b], John THOMAS [c], Mohanavelu KALATHE [d], Sreeraj V S [e], Venkat SUBRAMANIAM [f] and Jac Fredo AGASTINOSE RONICKOM [a]

[a] *School of Biomedical Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India*

[b] *Department of Electrical and Electronics Engineering, BITS Pilani, K.K.Birla Goa Campus, Goa, India*

[c] *Montreal Neurological Institute, McGill University, Canada*

[d] *Defence Bio-Engineering & Electro Medical Laboratory, DRDO, Ministry of Defence, Bangalore, India*

[e] *Clinical Research Centre for neuromodulation in psychiatry, NIMHANS, Bangalore, India*

[f] *Translational Psychiatry Laboratory, NIMHANS, Bangalore, India*

ORCiD ID: Jahanvi Jeswani https://orcid.org/0000-0001-8096-611X

Abstract. Electroencephalography (EEG) has recently gained popularity in user authentication systems since it is unique and less impacted by fraudulent interceptions. Although EEG is known to be sensitive to emotions, understanding the stability of brain responses to EEG-based authentication systems is challenging. In this study, we compared the effect of different emotion stimuli for the application in the EEG-based biometrics system (EBS). Initially, we pre-processed audio-visual evoked EEG potentials from the 'A Database for Emotion Analysis using Physiological Signals' (DEAP) dataset. A total of 21 time-domain and 33 frequency-domain features were extracted from the considered EEG signals in response to Low valence Low arousal (LVLA) and High valence low arousal (HVLA) stimuli. These features were fed as input to an XGBoost classifier to evaluate the performance and identify the significant features. The model performance was validated using leave-one-out cross-validation. The pipeline achieved high performance with multiclass accuracy of 80.97% and a binary-class accuracy of 99.41% with LVLA stimuli. In addition, it also achieved recall, precision and F-measure scores of 80.97%, 81.58% and 80.95%, respectively. For both the cases of LVLA and LVHA, skewness was the stand-out feature. We conclude that boring stimuli (negative experience) that fall under the LVLA category can elicit a more unique neuronal response than its counterpart the LVHA (positive experience). Thus, the proposed pipeline involving LVLA stimuli could be a potential authentication technique in security applications.

Keywords. Audio-visual evoked potential, Biometric authentication, Electroencephalography, Time and frequency features, XGBoost

[1] Corresponding Author: Praveen Kumar Govarthan, E-mail: praveenkumarg.bme21@itbhu.ac.in.

1. Introduction

Electroencephalography (EEG)-based biometrics is an emerging technology that utilizes brain signals to identify and authenticate individuals [1]. EEG signals can vary significantly depending on the user's mental task, as different areas of the cortex are activated and deactivated accordingly. The human mental state has a substantial impact on the brain's neuronal firing, making it highly sensitive to both external environmental stimulation and endogenous autonomous regulation [2]. Because of this, researchers consider EEG to be an ideal biometric system, as it can generate distinctive signals for each individual based on their brain's neural pathways and cognitive patterns [2]. Moreover, Electroencephalography based biometric systems (EBS) have the potential to surpass the limitations of conventional biometrics. The conventional biometric modalities have some security disadvantages such as the face, fingerprint and iris information that can be photographed, voice can be recorded, and handwriting can be mimicked. Moreover, individuals may lose or change their biometric characteristics such as finger or face in certain circumstances: changes may occur due to injury [3,4]. Further, EBS can provide a precise, non-intrusive, and dependable solution that can be used for a variety of applications, such as access control, medical diagnosis, and user authentication [5]. Despite current advances in EBS technology, there remain significant challenges in the selection of suitable paradigms for enrollment, universality, and user-friendliness in real-world EEG authentication scenarios. Out of these challenges, selecting the optimal paradigms is one of the most pronounced difficulties in EBS [2]. Many of the EEG-based biometric researches have used motor imagery [6], cognitive tasks [7], and response to visual stimuli [8] during enrollment in EBS identification and verification processes. However, each of these protocols has its own drawbacks. The performance of motor imagery and cognitive tasks is difficult and calls for extensive user training [8]. As suggested in [2], choosing a suitable induction paradigm will have a great impact on the recognition results. Recently, emotion-eliciting stimuli have been used for EBS applications, as a distinct brain neuronal firing pattern influenced by mood, stress, and mental state can be used as a possible biometric identifier [5,9]. Thus, in this study, we attempted to identify the optimal emotional eliciting stimuli required for developing credible EEG-based biometrics for usage in real-life scenarios.

2. Methods

The processing pipeline adopted in this study is shown in Figure 1. The EEG signals and subjective ratings of emotion experience from the 'A Database for Emotion Analysis using Physiological Signals' (DEAP) dataset were used in this investigation [10]. The EEG signals of 32 participants were recorded while they watched 40 emotion-eliciting music video clips. The data was collected across 32 channels with 8064 data points each. Participants rated each video in terms of the levels of arousal, valence, like/dislike, dominance and familiarity. Using subjective ratings, we segregated the video clips into four groups: High valence high arousal (HVHA), Low valence high arousal (LVHA), High valence low arousal (HVLA) and Low valence Low arousal (LVLA) according to the average valence and arousal ratings from all participants, keeping 5 as the threshold. The HVHA relates to amusing stimuli, LVHA can be scary stimuli, HVLA relates to relaxing stimuli, while LVLA is concerned with

boring stimuli. The preprocessed data from the DEAP dataset was used in this study, which had already been downsampled to 128Hz, EOG artifacts were removed, a 4.0-45.0Hz band-pass frequency filter was applied, data were averaged to the common reference, data was segmented into 60-second trials and a 3-second pre-trial baseline was removed [11].

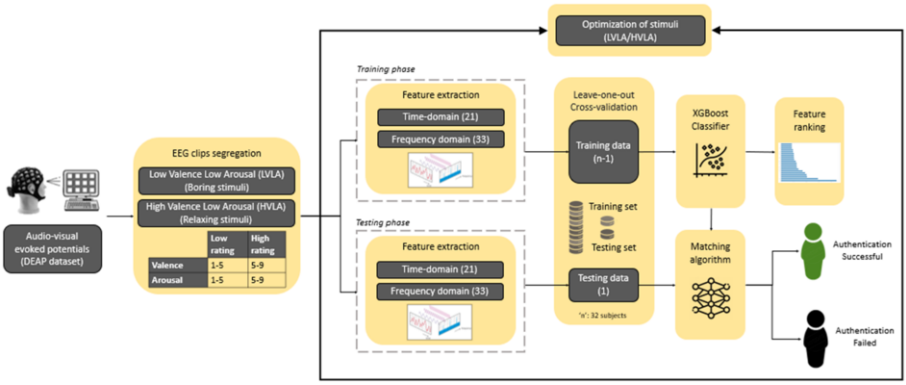


Figure 1. Proposed pipeline for EEG-based biometrics system

We extracted 21 time-domain and 33 frequency-domain features using data points for each user, each clip averaged out on all 32 channels resulting in a dataset with instances for user and clip combination [12]. The final dataset was split into a train and test set using leave-one-out cross-validation (LOOCV) where each instance was predicted, training on all other instances. The classification result was obtained using the eXtreme gradient boosting (XGBoost) classifier [12]. Two models were trained: one for LVLA, consisting of stimuli-based EEG segments corresponding to 11 video clips, and one for LVHA, consisting of stimuli-based EEG segments corresponding to 9 video clips. The performance parameters such as binary accuracy, multiclass accuracy, precision, recall and F-measure were calculated. We evaluated the model using two performance metrics for calculating classification accuracies: binary accuracy and multi-class accuracy. We have performed multiclass classification accuracy to remove bias induced due to true negatives and to get actual performance from the model. The feature ranking was then obtained using the feature importance method of the XGBoost algorithm [12].

3. Results and Discussion

Figure 2 shows the comparative performance of the XGBoost model built using LVLA and HVLA stimuli categories. It can be seen that multiclass accuracy for correctly identifying the user was 80.97% and 72.97% for LVLA and HVLA, respectively. The LVLA stimuli-based EEG segments yielded higher average classification accuracy than HVLA. It indicates that LVLA stimuli were better at eliciting the emotion required for EEG-based biometrics application as compared to the HVLA. F-measure, recall and precision were also high using LVLA stimuli-based EEG segments. However, the binary class accuracy is similar in both cases.

The feature importance plots corresponding to the XGBoost model built using LVLA and HVLA stimuli are shown in Figure 3a and Figure 3b respectively. We illustrated

the top 10 most relevant features out of 54 features considered in the analysis. Skewness, Higuchi's fractal dimension and sample entropy are the best features for authenticating individuals using EEG signals, for both LVLA and HVLA.

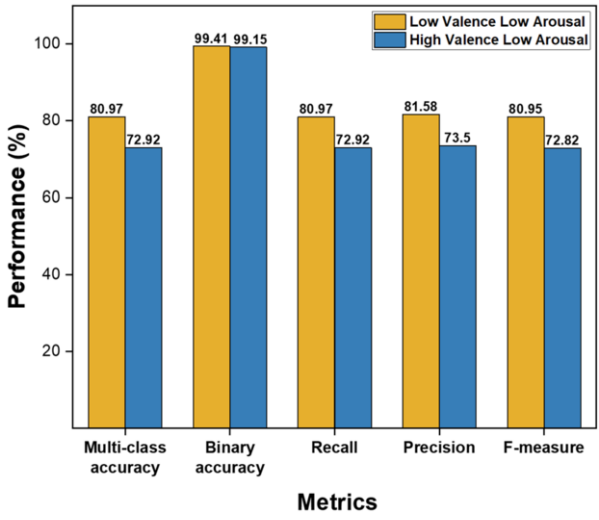


Figure 2. Comparison of classifier performance in two different stimuli: LVLA and HVLA

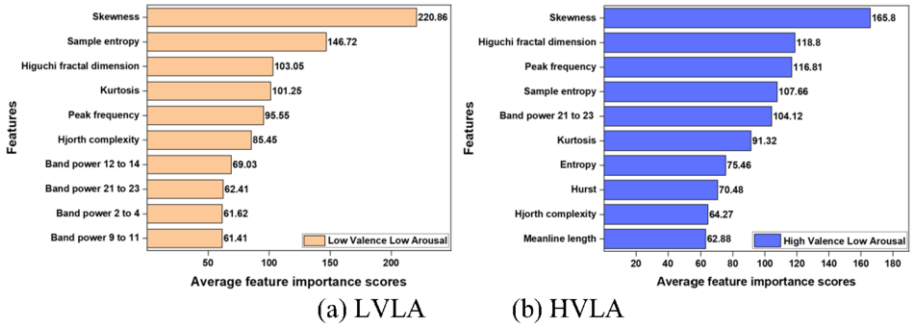


Figure 3. Average feature importance scores (in descending order) achieved in each stimulus

4. Limitations and Future work

We have compared only LVLA and HVLA audio-visual stimuli to find the optimal stimuli for EEG-based biometrics; however, in the future, we could extend this study to other stimuli such as LVHA and HVHA to obtain more comprehensive results. Moreover, while this study has only used the XGBoost model, more complex models or other ensemble learning methods could be employed to obtain better predictive performance. Additionally, more sophisticated deep learning architectures could also be utilized to improve the classifier performance in authentication. It is essential to compare the performance of the model on multiple datasets and real-time persons from different demographics to increase the sample size and assess its generalizability.

Additionally, we have only utilized time and frequency domain features for authentication, however, the use of time-frequency domain features could provide intricate details that could improve the efficiency of the matching algorithms in the authentication. Furthermore, we were unable to find any exact literature that implements a similar approach using the DEAP dataset for EEG-based authentication. However, we will compare our results with other EEG-based authentication studies in the future.

5. Conclusions

The results of our study show that authentication using audio-visual evoked EEG signals can be performed successfully. Our model was validated with LOOCV, a low-bias method and achieved the highest multi-class classification and binary classification accuracies of 80.97% and 99.41%, respectively, with LVLA stimuli. Our analysis proved that the LVLA is a better candidate than HVLA for EEG-based biometrics applications. Further, the identification of best-performing features was analyzed, and the skewness was found to be the standout feature in both LVLA and HVLA stimuli. Our results suggest that with further development and refinement, EEG-based biometrics systems have the potential to revolutionize the way of biometric authentication.

References

[1] Bidgoly AJ, Bidgoly HJ, Arezoumand Z. Towards a universal and privacy preserving EEG-based authentication system. Scientific Reports. 2022 Feb 15;12(1):1-2.
[2] Zhang S, Sun L, Mao X, Hu C, Liu P. Review on EEG-Based Authentication Technology. Computational Intelligence and Neuroscience. 2021 Dec 24;2021.
[3] Blanco-Gonzalo R, Lunerti C, Sanchez-Reillo R, Guest RM. Biometrics: Accessibility challenge or opportunity? PloS one. 2018 Mar 22;13(3):e0194111.
[4] Yang W, Wang S, Sahri NM, Karie NM, Ahmed M, Valli C. Biometrics for internet-of-things security: A review. Sensors. 2021 Sep 14;21(18):6163.
[5] Khalil R, Arasteh A, Sarkar AK. EEG based biometrics using emotional stimulation data. In2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 2017 Dec 21 (pp. 246-249). IEEE.
[6] Nguyen P, Tran D, Huang X, Ma W. Motor imagery EEG-based person verification. International work-conference on artificial neural networks 2013 Jun 12 (pp. 430-438). Springer, Berlin, Heidelberg.
[7] Palaniappan R. Two-stage biometric authentication method using thought activity brain waves. International journal of neural systems. 2008 Feb;18(01):59-66.
[8] Brigham K, Kumar BV. Subject identification from electroencephalogram (EEG) signals during imagined speech. In2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) 2010 Sep 27 (pp. 1-8). IEEE.
[9] Wilaiprasitporn T, Ditthapron A, Matchaparn K, Tongbuasirilai T, Banluesombatkul N, Chuang-suwanich E. Affective EEG-based person identification using the deep learning approach. IEEE Transactions on Cognitive and Developmental Systems. 2019 Jun 25;12(3):486-96.
[10] Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing. 2011 Jun 9;3(1):18-31.
[11] Yang Y, Wu Q, Fu Y, Chen X. Continuous convolutional neural network with 3D input for EEG-based emotion recognition. International Conference on Neural Information Processing 2018 Dec 13 (pp. 433-443). Springer, Cham.
[12] Abirami S, Thomas J, Yuvaraj R, Agastinose Ronickom JF. A Comparative Study on EEG Features for Neonatal Seizure Detection. In Biomedical Signals Based Computer-Aided Diagnosis for Neurological Disorders 2022 (pp. 43-64). Springer, Cham.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230115

Medical Secretaries' Registration Work in the Data-Driven Healthcare Era

Casper KNUDSEN[a,1] and Pernille BERTELSEN[a]
[a] *Aalborg University, Department of Planning, Denmark*
ORCiD ID: Casper Knudsen https://orcid.org/0000-0002-9491-801X,
Pernille Scholdan Bertelsen https://orcid.org/0000-0003-2276-5496

Abstract. Through a qualitative study in six hospital departments in the Northern Region of Denmark, this article aims to shed light on how a non-clinical group, medical secretaries, supports clinical-administrative documentation as they translate between the clinical and administrative domains. This article shows how this demands context-sensitive knowledge and skills acquired through deep engagement with the full scope of clinical-administrative work at the department level. We argue that, given the increasing ambitions for secondary uses of healthcare data, specific clinical-administrative competencies beyond those of clinicians are increasingly necessary in the skillmix in hospitals.

Keywords. registration practices, medical secretaries, EHR, electronic health record, clinical-administrative documentation, data work, data quality

1. Introduction

Healthcare in Scandinavia and elsewhere is becoming increasingly data-driven, and governments and healthcare organizations are investing heavily in health information technology, aiming to harvest the fruits of secondary data use [1]. As the ambitions and prospects for data use increase, so does the demand for more and increasingly accurate and complete clinical-administrative datasets. This development persistently puts healthcare professionals under pressure to deliver quality documentation [2,3] and challenges the organization of those tasks.

This is manifestly emphasized by the issue of clinician burnout, which has been extensively reported for the past decade. Burnout in healthcare is now being linked to digitalization and, in particular, electronic health record (EHR) use, as time spent by clinicians on documentation is rising [4] but also leading to the emergence of new occupations such as medical scribes [3]. While the "EHR burnout" debate is still less apparent in the Danish research context, where the study behind this article took place, discussion continues among practitioners [5] and in the public political debate. Here, the issue of resource allocation and the burden of physician documentation has led to discussions about enlisting the help of non-clinical professionals, such as medical secretaries, in managing the documentation and registration workload.

[1] Corresponding Author: Casper Knudsen, E-mail: Casper@plan.aau.dk

This paper follows a recent stream of socio-technical research investigating on-the-ground data work in healthcare [6], as new ambitions for becoming data-driven lead to new professions [3,7] and encourage existing professions to take on new tasks [8–11]. The ambition is to illuminate the work tasks involved, as "data do not sit in ready repositories, fully formed, and easily harvestable. Data must be created through various forms of situated work" [6].

While large-scale data involves the work of many people and professions [12], less prominent professions in health care remain under-researched. This paper positions itself in the body of literature aiming to understand and highlight the – often invisible [13] – work of health administrative personnel in hospitals, such as secretaries, in Norway [14] and Denmark [8,10,15–17]. These studies emphasize the significance of health administrative personnel in creating and sustaining cohesion in hospitals. In Danish public hospitals, medical secretaries work in decentralized units at department and ward levels, close to clinical practice and within a specific clinical specialty in which they build up their situated skills and knowledge [16]. Here, medical secretaries perform a wide variety of clinical-administrative work tasks, ranging from specialized data registration and coding to reception work, staff plans, department finance, management support, etc.[17]. Studies emphasize their decentralized location /distribution as a key factor that enables them to "often act as the organizational 'glue' or connecting thread between health care professionals at the hospital" [17] and be positioned at the intersection of clinical and administrative work [15].

Empirically, we investigated the data work of medical secretaries in hospital departments to understand how and why their data work contributes to keeping data complete and accurate. We thereby aimed to understand the skills involved in their registration practices and how administrative work is interwoven into the local medical context.

2. Methods

This article reports on an empirical, qualitative multi-site study of the data work of medical secretaries in the Northern Region of Denmark. Interviews (n = 9, 1–1.5 hours each) and observations (n = 9, total 27 hours) of medical secretaries responsible for registration across 6 somatic departments in 3 hospitals were conducted. Additionally, interviews (n = 2) with staff from the central finance department and the heads of medical secretaries (n = 3) were included in the analysis. The observation and the interviews were carried out in January–April 2022 and November–December 2022 in the same departments, with one department unable to participate in the last round of interviews. All the participants received and signed an informed consent form.

To understand the data work of medical secretaries, the interviews and observations focused on the content, knowledge and competencies, organization and collaboration, and systems and tools related to their work. All interviews were fully transcribed. The transcripts and field notes from the observations were coded and categorized. The findings reported in this paper represent emerging themes in the dataset, as the analysis is continuing.

3. Results

The results of our study demonstrate the crucial role played by medical secretaries in ensuring the accuracy and completeness of clinical documentation in EHRs and other information systems.

While physicians have the formal responsibility for meaningful and adequate clinical documentation, the practical work of registration and data quality management at department level is mainly delegated to medical secretaries[16]. The registration of patient trajectory data in the EHR and other information systems is a key task in the work portfolio of medical secretaries, usually undertaken as an integrated part of the workflow from patient referral to diagnosis, treatment, discharge, and follow-up. Medical secretaries transcribe notes from physicians, assign correct labels and markers to the patient trajectory, ensure that timestamps are correct, and conduct ongoing monitoring, ensuring that the department's body of registration data is as complete, correct, and aligned with the requirements of the registration logic as possible.

In the departments' division of work, the role of the medical secretaries is to translate between the clinical and administrative domains, which is done through registration of the data based on the physicians' clinical notes, extending the clinical documentation with administrative codes specifying the type of intervention, the reasons for waiting time, etc. Across all the departments interviewed, the medical secretaries explain how, based on their experience of working in that department's medical field, they "fill in the blanks" when information is missing in the documentation from the physician, e.g., when a diagnosis has not been recorded at the first consultation, as required by the registration guidelines. When describing their repair work in relation to patient data, our informants often refer to it in terms of "tidying up" or even requiring "a registration brain" (Medical secretary 2) or "a healthy bit of OCD" (Medical secretary 3), highlighting the professional virtue of attention to detail in the administrative work as essential for applying and managing correct decentralized registration, as requirements for registration increase in volume and complexity.

Beyond the technical skill of applying a trajectory-oriented registration regime, substantial knowledge of the intricacies of registration within the EHR and other information systems and their – often not seamless – interplay in the department's concrete organizational and clinical setting is essential for the secretaries to achieve accurate data by ensuring that interventions set up in the EHR by physicians are linked to the correct trajectory, correctly classified to count against department activity, and correctly time stamped. For example, the latter often results in teleconsultations being erroneously registered as taking hours or even days, when automatic timestamping of events collides with work practices of coding after the fact rather than in real time. A key aspect in the role is translating between the administrative and clinical aspects of documentation: *"Basically, they [clinicians] don't understand this kind of system at all. This whole thing about it mattering whether you make **this** registration at **this** specific time – that it needs to be within this contact – well, it's like a foreign language to them, totally incomprehensible"* (Medical Secretary 1). Our informants clearly state that their perception is that documentation requirements exceed physicians' interest and skills, as demands for registration are often linked to secondary uses of data rather than immediate clinical use. As another informant frames it: *"They [the physicians] do not at all have that basic knowledge, which secretaries are brought up with, in terms of how important data is […] They just want to be able to click something to make the system accept [it], so they can move on. Therefore, they will not have the same level of attention to detail*

as the secretaries. Instead, the secretaries will be spending much more time on quality assurance – and hence the data you get at first will not be valid" (Head of Medical Secretaries 1).

While registration work in the regular workflow of the department relies on context sensitivity obtained through practical, daily experience, error correction is done by medical secretaries through automated lists of logical errors [16], our study finds, requiring a significantly higher level of expertise: *"A prerequisite for that is also being able to understand the system, knowing the connections in the entire department as well. It's difficult just to come from the outside – you have to be familiar with the department's flow of different things"* (Secretary in Cardiology).

From our data, it is clear that the task of correcting errors is more complex and in the sites involved undertaken by medical secretaries with considerable (often 10+ years) experience.

4. Discussion and Conclusion

Hogle [1] asserts that "big data and the infrastructures instilled to support it are creating new forms of value and reordering relationships as distinctions between medical and non-medical data and research and care are blurred." Organization of the work presented in this paper can be seen as a consequence of the blurring line between medical and non-medical data. While clinical documentation is clearly the physician's responsibility, the administrative requirements of clinical documentation go beyond the clinical domain to the extent that some researchers question whether requirements for secondary data purposes take precedence over the primary clinical purposes in the design of EHRs [2]. In the example of our case, a trajectory-oriented registration regime in which all activities are registered according to their relation to a specific illness trajectory increases the socio-technical complexity in the documentation workflow, as this logic often does not match the clinical intervention as perceived by the clinician. A distinct clinical-administrative perspective is required.

The increasing complexity of documentation creates a need for translation between the clinical and administrative domain [7], which in the case presented here is undertaken by medical secretaries. The work to commit registration to the clinical documentation and ensure alignment with the registration regime at hand is essential for harvesting the potential of initiatives toward data-intensive resourcing for management, political oversight, and research.

Physician burnout and the strain of an increasing documentation burden on clinicians are indeed serious issues; however, we argue that the discourse of simply relieving physicians of documentation tasks risks underestimating the knowledge and competencies necessary to secure valid data in the clinical-administrative domain. Our research indicates that effective planning for the healthcare skill mix must consider the complexities of on-the-ground data work following from ambitious secondary data initiatives. Additionally, our research suggests a need for looking beyond the major, most influential professions in discussing the design of systems, practice, and organization for secondary data use, acknowledging and actively engaging the distinct skills and virtues required in complex clinical-administrative work in the context of data-intensive resourcing.

References

[1] Hogle LF. Data-intensive resourcing in healthcare. BioSocieties. 2016;11:372–93. doi:10.1057/s41292-016-0004-5.

[2] Green S, Hillersdal L, Holt J, Hoeyer K, Wadmann S. The practical ethics of repurposing health data: how to acknowledge invisible data work and the need for prioritization. Med Health Care and Philos. 2022. doi:10.1007/s11019-022-10128-6.

[3] Bossen C, Chen Y, Pine KH. The emergence of new data work occupations in healthcare: the case of medical scribes. Int J Med Inform. (Shannon, Ireland) 2019;123:76–83. doi:10.1016/j.ijmedinf.2019.01.001.

[4] Kruse CS, Mileski M, Dray G, Johnson Z, Shaw C, Shirodkar H. Physician burnout and the electronic health record leading up to and during the first year of COVID-19: systematic review. J Med Internet Res. 2022;24:e36200. doi:10.2196/36200.

[5] Bentzon N, Rosenberg J, Bansler J, Søgaard J, Dahlerup B, Haarder B. Destruktiv digitalisering: en debatbog om Sundhedsplatformen 2016-2021. Copenhagen: FADL; 2021.

[6] Bossen C, Pine KH, Cabitza F, Ellingsen G, Piras EM. Data work in healthcare: an introduction. Health Informatics J. 2019;25:465–74. doi:10.1177/1460458219864730.

[7] Pine KH, Bossen C. Good organizational reasons for better medical records: the data work of clinical documentation integrity specialists. Big Data Soc. 2020;7. doi:10.1177/2053951720965616.

[8] Bossen C, Jensen LG, Udsen FW. Boundary-object trimming: on the invisibility of medical secretaries' care of records in healthcare infrastructures. Computer Supported Cooperative Work (CSCW) 2014;23:75–110.

[9] Islind AS, Lindroth T, Lundin J, Steineck G. Shift in translations: data work with patient-generated health data in clinical practice. Health Informatics J. 2019;25:577–86. doi:10.1177/1460458219833097.

[10] Knudsen C, Bertelsen PS. Data work in health care: the case of medical secretaries. Studies in Health Technology and Informatics 2021;286:60–4. doi:10.3233/SHTI210638.

[11] Bjørnstad C, Ellingsen G. Data work: a condition for integrations in health care. Health Informatics J. 2019;25:526–35. doi:10.1177/1460458219833114.

[12] Holten NM, Pine KH, Nielsen TR, Bossen C. Who does the work of data? Interactions 2020. doi:10.1145/3386389.

[13] Star S, Strauss A. Layers of silence, arenas of voice: the ecology of visible and invisible work. Computer Supported Cooperative Work 1999;8:9–30. doi:10.1023/A:1008651105359.

[14] Håland E, Melby L. Coding for quality? Accountability work in standardised cancer patient pathways (CPPs). Health (London, England : 1997) 2021:13634593211013882. doi:10.1177/13634593211013882.

[15] Møller NLH, Vikkelsø S. The clinical work of secretaries: exploring the intersection of administrative and clinical work in the diagnosing process. In: Dugdale, J., Masclet, C., Grasso, M., Boujut, JF., Hassanaly, P. (eds) From Research to Practice in the Design of Cooperative Systems: Results and Open Challenges. London. Springer; 2012, p. 33–47. doi:10.1007/978-1-4471-4093-1_3.

[16] Knudsen C, Bertelsen P. Maintaining data quality at the hospital department level – the data work of medical secretaries. Scandinavian Conference on Health Informatics 2022:159–65. doi:10.3384/ecp187027.

[17] Bertelsen P, Nøhr C. The work practice of medical secretaries and the implementation of electronic health records in Denmark. Health Inf Manag J. 2005;34.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230116

HEIDA: Software Examples for Rapid Introduction of Homomorphic Encryption for Privacy Preservation of Health Data

Rickard BRÄNNVALL[1], Henrik FORSGREN and Helena LINGE
RISE Research Institutes of Sweden

Abstract. Adequate privacy protection is crucial for implementing modern AI algorithms in medicine. With Fully Homomorphic Encryption (FHE), a party without access to the secret key can perform calculations and advanced analytics on encrypted data without taking part of either the input data or the results. FHE can therefore work as an enabler for situations where computations are carried out by parties that are denied plain text access to sensitive data. It is a scenario often found with digital services that process personal health-related data or medical data originating from a healthcare provider, for example, when the service is delivered by a third-party service provider located in the cloud. There are practical challenges to be aware of when working with FHE. The current work aims to improve accessibility and reduce barriers to entry by providing code examples and recommendations to aid developers working with health data in developing FHE-based applications. HEIDA is available on the GitHub repository: https://github.com/rickardbrannvall/HEIDA.

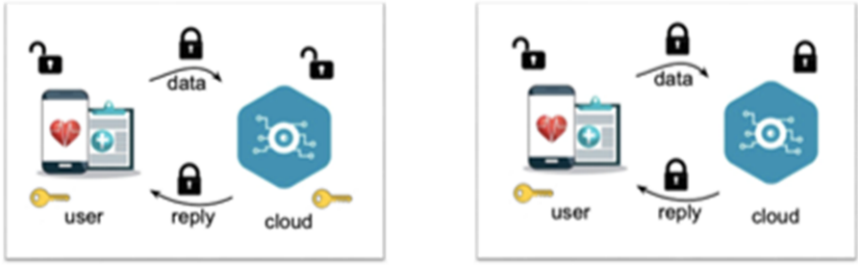
Keywords. Artificial Intelligence, GDPR, Sensitive Data, Privacy Preservation

1. Introduction

The application of AI and machine learning has made remarkable advances in healthcare in the last decade and holds promise to revolutionize the field by improving diagnosis, treatment, and prevention of diseases, e.g., in applications of personalized medicine [1] or medical image analysis [2]. The use of AI in healthcare applications raises concerns as it involves processing sensitive personal information. Privacy is a fundamental human right protected by privacy laws and regulations, e.g., GDPR in the EU and HIPAA in the USA. Adequate privacy protection builds trust and facilitates the sharing of personal health information. This can enable the development of more comprehensive and integrated healthcare services that improve patient care and outcomes, as well as develop commercial opportunities. Several alternative approaches have been proposed, e.g., cryptographic techniques, differential privacy, and federated learning [3]. We aim to show the feasibility of homomorphic encryption for this end.

The remainder of this section briefly reviews Fully Homomorphic Encryption (FHE) and challenges it poses to developers. Following sections discuss methods, present results for the library, and then conclude with lessons learned.

[1] Corresponding Author: Rickard Brännvall, E-mail: rickard.brannvall@ri.se



(a) conventional. (b) homomorphic.

Figure 1. Comparison of alternative implementations of a digital health service.

1.1. Background

Homomorphic encryption was first proposed as a mathematical challenge in the late 1970s by Rivest, but it remained unsolved until 2009, when Gentry presented the first feasible scheme. Progress had been made in the interim years with schemes that offered partial homomorphic encryption, e.g., for either multiplication or addition, but not both. Gentry's construct is the first Fully Homomorphic Encryption (FHE) scheme that supports arbitrary depth computational circuits through the so-called bootstrap. After this breakthrough, a decade of intense research followed, and recently practical solutions have emerged[2]. This work uses the TFHE scheme by Chillotti et al. [5] as implemented in the Concrete [6] library for the Rust programming language.

Data sharing under FHE. We use advanced encryption technology daily: web browsers, car-key fobs, and bank transfers all rely on conventional encryption that protects sensitive data while it is stored or transferred. However, conventional encryption technologies require data to be decrypted before it is processed, as illustrated in Figure 1a for a digital health service hosted in the cloud. Note in the (left) figure how the cloud must have access to the secret decryption key. Figure 1b (right panel) depicts the same service delivered under homomorphic encryption where the cloud does not need the secret key as it can directly process the encrypted data. The cloud returns the results in an encrypted reply without ever having had plain text access to the original data or any results. Besides the benefits of oblivious computation, FHE also prevents unintended secondary use of data – even if sold further, encrypted data cannot be meaningfully used for other purposes. Future privacy remain as long as the secret key is kept safe.

Some notations. The method is named after the mathematical notion of homomorphism, which is used to signify that elements of one set are transformed into elements of a second set while maintaining the relationships between the individual elements in each of the sets; for example, for the arithmetic operations: $E(x+y) = E(x) + E(y)$ and $E(xy) = E(x)E(y)$, whereby $E(x)$ we mean applying the encryption operation to plaintext variable x. It is a common convention to use the word *plain text* for unencrypted data and *cipher text* for encrypted data. Decryption, which maps a cipher text to its corresponding plain text element, is possible only for the secret key holder. Modern FHE schemes are considered unbreakable under very strong cryptographic guarantees and held to be resilient even against hypothetical quantum computer-based attacks [7].

Programming for FHE is complex and may require both cryptography and numerical computation expertise. To ameliorate this, FHE libraries like the Simple Encrypted

[2] Many references in this section are omitted for brevity. Please consult [4] for a comprehensive review.

Arithmetic Library (SEAL) [8] or the Fast Fully Homomorphic Encryption over the Torus (TFHE) [6] implement the underlying cryptographic operations and expose a higher-level API. Viand et al. [9] recently surveyed different libraries and compilers to compare FHE tools and sum up state-of-the-art. Examples of FHE applications, including medical, are provided by Optalysys [10], OpenFHE [11], or Zama [12].

Programming paradigm. The computation model used for FHE provides a challenge for algorithm implementation. Under the standard programming paradigm, one can use flow control operations to arbitrarily break loops when a specific condition is observed (while-loops) or conditionally select which code branch to execute (if-then-else statements). Program flow for FHE computations cannot, by necessity, depend on the data without violating privacy guarantees. Otherwise, an attacker with a clock can get information about encrypted variables by following the program flow of branching code through a so-called timing attack.

Noise management. Numerical precision is generally of little concern when programming in a modern high-level language like Python; with FHE programming, it instead becomes the central concern. By construction, the homomorphic encryption introduces noise into the data. This noise grows for every operation, potentially rendering the results of large computations meaningless even after decryption. Some strategies to manage this problem are bootstrapping, level constraints, and client-side refresh. *Bootstrapping* is possible when the scheme has sufficient capacity to execute its own decryption and encryption circuits under encryption. This refreshes the cipher text by removing some but not all cryptographic noise. In a *leveled approach*, one carefully has to design the algorithm and limit the number of operations such that the results are obtained before bootstrapping becomes necessary. Practically, only simple algorithms without iteration are feasible. The depth of the circuit depends on the parameter choice for the encryption scheme, which we will refer to as key size. *Client-side refresh* combines a leveled approach with roundtrips back to the client. This replaces the costly server-side bootstrap operation with a decryption re-encryption on the client side, which similarly cleans the data of cryptographic noise. It comes at the cost of additional communication overhead and potentially reveals intermediate results of the computation to the client. It is suitable when there naturally are iterations of data transfer between parties, such as remote monitoring and control applications or federated learning.

Computational overhead. The privacy benefits FHE offers come at a price, as it is generally computationally very demanding. It can show orders of magnitude slower execution compared to conventional computation, even in leveled approaches that do not use bootstrapping (which is slow because it executes a complex circuit). It also places greater demands on storage and memory as data files and keys generally have large sizes. Some FHE schemes only support arithmetic operations (addition, multiplication, subtraction) and are hence limited to polynomial function evaluation, but the TFHE scheme selected for this work [6] can approximate general functions through the functional bootstrap [5].

2. Method

HEIDA is an open-source library with examples of health data analysis under homomorphic encryption. It is built in the Concrete library for FHE on the Torus by Zama [6] and designed to be easily imported and used for model performance assessment. HEIDA includes standardized helper functions for key generation, data encryption, and

some higher-level mathematical operations. It is hosted, supported, and version-controlled in a repository on GitHub[3] under the open-source BSD 3-Clear License.

We tested executing with and without encryption to validate the results and compare computational performance. Different key sizes and other parameters were tested and summarized in tables that can assist with the planning and design of data structures and code for a specific problem. To facilitate the use of the library code, we provide worked-through examples of how to implement simple analysis of health data under FHE.

The code examples were chosen to illustrate common machine learning techniques and to show how to overcome technical challenges, such as conditional statements of branching trees or iteration over time series. Only open data sets that did not include personal information were used to avoid privacy concerns with the open-source release.

Example 1: Logistic regression for coronary disease. A logistic regression model trained to estimate a patient's risk of future coronary heart disease based on data about demographic, behavioral, and medical factors. It illustrates a simple machine-learning prediction method implemented through a leveled approach (without bootstrapping).

Example 2: Cardiac disease risk assessment. This is for a similar application as example 1 but built on different input data and a different method of inference. It is included to illustrate how a decision-tree structure can be handled within the constrained programming paradigm that doesn't permit conditional execution of branching code. It also gives an example of parallel execution to make FHE applications run faster.

Example 3: Diabetes self-care analysis. Continuous blood glucose measurements are analyzed under FHE to provide risk scores and gamified advice. This example is based on a solution [13] that was selected as the winner of the 2021-22 Vinter innovation competition arranged by the Swedish innovation agency Vinnova. It showcases time-series analysis with scalable performance and how to implement general non-polynomial functions by applying the functional bootstrapping method [5].

Example 4: Activity monitoring by a neural network. We assume that encrypted personal data, such as blood glucose measurements, carbohydrate intake, physical activity, and insulin administration, to be analyzed by a neural network.

Example 5: Federated learning with encrypted aggregation. A ResNet-18 deep learning model is trained by transfer learning for the classification of dermatoscopic images. This is an example of how to realize gradient-based training of a deep-learning model by a distributed algorithm taking a leveled approach in each aggregation sub-steps which is iterated over many training epochs. It is possible because federated learning inherently builds on data communication round-trips between multiple parties.

3. Discussion

FHE allows computation without exposing data, but it can be computationally very expensive. It is most suitable when one party (a) owns sensitive data that must be processed by a party (b) that is not privy to plain text access to this data. For example, when (b) possesses a superior proprietary algorithm that must be used but cannot be shared with (a). This is illustrated in the case of Electronic Health Records (EHR) in examples 1 and 2; and for what could be considered Personal Health Records (PHR) in examples 3 and 4. Instead of using a proprietary algorithm, it can also be that the tasks consist of combining a second confidential data set that (b) owns but cannot share with

[3] https://github.com/rickardbrannvall/HEIDA

(a). This is the case in example 5, which also illustrates the intended secondary use of data. Note that modern FHE schemes protect data against unintended future use due to their quantum computer resilience, which is especially important for EHR and PHR.

As an alternative, secure multiparty computation can be used to jointly compute any function, but it can have a significant communication overhead and furthermore requires complex coordination between parties (at least compared to the simplicity of FHE cloud deployment in Figure 1b). Differential privacy protects data by adding noise and can be a good complement to federated learning, but it can adversely affect model accuracy.

Lessons Learned. We consider FHE to be most appropriate for high-value applications of moderate complexity at the level of the examples given in Section 2. Large-scale deep learning models, such as transformer language models, are infeasible for implementation under FHE, at least at the time of the writing of this article. The same would count for high-velocity data applications like online video analysis.

On-going work utilizing the library includes collaboration with two regional healthcare providers on personal data protection, particularly with an aim to explore synergies with other advanced privacy-preserving technologies such as federated learning. Currently, we are also investigating new neural network architectures that execute more efficiently under encryption, particularly with an eye to Natural Language Processing.

Conclusions. Homomorphic encryption is a technology that enables cloud services with strong privacy. This work presents open-source[3] software examples intended to facilitate the use of homomorphic encryption for privacy protection in the medical data sector. Future directions include providing Python bindings as well as building software examples for recurrent neural networks.

References

[1] Schork NJ. Artificial intelligence and personalized medicine. In: Precision medicine in Cancer therapy. Springer; 2019. p. 265-83.
[2] Litjens G, et al. A survey on deep learning in medical image analysis. Med Image Analysis. 2017 dec.
[3] Torkzadehmahani R, et al. Privacy-preserving AI techniques in biomedicine. Methods of Information in Medicine. 2022.
[4] Halevi S. Homomorphic Encryption. In: Tutorials on the Foundations of Cryptography. Springer International Publishing; 2017. p. 219-76.
[5] Chillotti I, et al. TFHE: Fast Fully Homomorphic Encryption Over the Torus. Journal of Cryptology. 2019 apr;33(1):34-91.
[6] Chillotti I, et al. CONCRETE: Concrete Operates oN Ciphertexts Rapidly by Extending TFHE. WAHC 2020–8th Workshop on Encrypted Computing & Applied Homomorphic Cryptography; 2020.
[7] Augot D, et al. Initial recommendations of long-term secure post-quantum systems; 2015.
[8] Microsoft SEAL (release 3.6); 2020. Microsoft Research, Redmond, WA. Accessed 2023-03-05. https://github.com/Microsoft/SEAL
[9] Viand A, Jattke P, Hithnawi A. SoK: Fully Homomorphic Encryption Compilers. In: 2021 IEEE Symposium on Security and Privacy (SP). IEEE; 2021.
[10] Optalysys;. Accessed 2023-03-05. https://github.com/jw-optalysys
[11] OpenFHE;. Accessed 2023-03-05. https://www.openfhe.org/publications
[12] Zama;. Accessed 2023-03-05. https://github.com/zama-ai
[13] Brännvall R, et al. Homomorphic encryption enables private data sharing for digital health: winning entry to the Vinnova innovation competition Vinter 2021-22. In: Proc. Swedish AI Symposium; 2022.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230117

Identifying Relevant FHIR Elements for Data Quality Assessment in the German Core Data Set

Christian DRAEGER[a,1], Erik TUTE[b], Carsten Oliver SCHMIDT[c], Dagmar WALTEMATH[d], Martin BOEKER[e], Alfred WINTER[a] and Matthias LÖBE[a]
[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany*
[b] *Peter L. Reichertz Institute for Medical Informatics (PLRI), Hannover Medical School, Germany*
[c] *Institute for Community Medicine, University Medicine Greifswald, Germany*
[d] *Medical Informatics Laboratory and Data Integration Center, University Medicine Greifswald, Germany*
[e] *Institute of Artificial Intelligence and Informatics in Medicine, Technical University of Munich, Germany*
ORCiD: Christian Draeger 0000-0001-8835-4548, Erik Tute 0000-0002-9176-3913, Carsten Oliver Schmidt 0000-0001-5266-9396, Dagmar Waltemath 0000-0002-5886-5563, Martin Boeker 0000-0003-2972-2042, Alfred Winter 0000-0003-0179-954X, Matthias Löbe 0000-0002-2344-0426

Abstract. The German Medical Informatics Initiative makes clinical routine data available for biomedical research. In total, 37 university hospitals have set up so-called data integration centers to facilitate this data reuse. A standardized set of HL7 FHIR profiles ("MII Core Data Set") defines the common data model across all centers. Regular Projectathons ensure continuous evaluation of the implemented data sharing processes on artificial and real-world clinical use cases. In this context, FHIR continues to rise in popularity for exchanging patient care data. As reusing data from patient care in clinical research requires high trust in the data, data quality assessments are a key point of concern in the data sharing process. To support the setup of data quality assessments within data integration centers, we suggest a process for finding elements of interest from FHIR profiles. We focus on the specific data quality measures defined by Kahn et al.

Keywords. Data Quality, HL7 FHIR, EHRs, Secondary Use

1. Introduction

The Medical Informatics Initiative (MII) makes healthcare data from German hospitals accessible and interoperable for research. Therefore, EHR data is extracted from the source systems in all German university hospitals, transformed into a common data

[1] Corresponding Author: Christian Draeger, Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany; E-mail: christian.draeger@imise.uni-leipzig.de

model, and made available via the data integration centers (DIC) [1]. The MII agreed to use HL7 FHIR as their interoperability standard at all 37 DIC sites. A common data model for the so-called MII Core Data Set (CDS) has been defined using FHIR profiles, which allow data to be represented and captured in a structured way. The CDS contains, among others, the profiles *Condition, Observation, Patient,* and *Encounter.* Each profile defines a set of relevant elements, described and published in the FHIR registry *simplifier*[2]. Researchers can request CDS elements from a central portal[3] to receive data from the DICs for research purposes. If a request passes the legal and ethical requirements, that DIC will provide the requested elements in FHIR bundles.

Projectathons are conducted to monitor the progress of the MII by evaluating the data sharing processes and infrastructures. However, the secondary use of healthcare data in real-world settings requires more consideration of data quality (DQ) [2]. Data collected for routine care is not automatically suitable for research purposes in terms of quality and granularity. While quantitative measures can give an indication of the data sets' suitability, a detailed data quality assessment (DQA) is required in the context of a specific use case and for each data provision. Compared with the data collections in clinical trials and epidemiological datasets, it is impossible to influence data collection in routine clinical practice. For example, data collections in clinical routine do not follow a specified sampling frame and standardized examination protocol. Therefore, available tools for data quality assessment in clinical and epidemiological studies are only partially suitable for this application scenario. Additionally, DQA tools developed in the MII [3,4] do not work directly on the FHIR bundles provided by the DIC.

Given this lack of DQA tools for the structured, complex FHIR data stored in the DICs, we (1) analyzed the FHIR profiles established in the CDS in an effort to identify elements with the highest impact toward a comprehensive DQ analysis, and (2) aligned them with the Kahn et al. [5] terminology framework. The Kahn framework was chosen as the base terminology for DQ discussions within the MII. Following the MII decision, we used the Kahn framework to identify relevant FHIR elements for each of Kahn's definitions, taking ongoing implementations for current Projectathons into account.

2. Methods

The recent 6th Projectathon set out to evaluate the data-sharing process, starting with a data request and finishing with a distribution of the required CDS elements as described in the introduction. With this goal in mind, the research question was formulated as: which value of the laboratory parameter NT-proBNP is a suitable marker for the diagnosis of cardiological diseases such as atrial fibrillation, taking age and gender into account? Apart from general patient information, only one laboratory measurement was used.

We first analyzed the CDS with regard to relevant elements for a DQA and then implemented DQ measures for all elements requested in the 6th Projectathon. Kahn et al. [5] defined a comprehensive DQ framework for the secondary use of EHR data. The framework is defined using three categories of data quality. The first category, *Conformance*, "focuses on DQ features that describe the compliance of the representation of data against internal or external formatting, relational, or computational

[2] https://simplifier.net/organization/koordinationsstellemii

[3] https://forschen-fuer-gesundheit.de/

definitions" [5]. The second category, *Completeness*, "focuses on features that describe the frequencies of data attributes present in a data set without reference to data values". The third category, *Plausibility*, "focuses on features that describe the believability or truthfulness of data values". These three main categories are further divided into subcategories, and lastly into definitions represented by one letter. The framework differentiates between the internal context, *Verification*, which considers "how data values match expectations with respect to metadata constraints, system assumptions, and local knowledge", and the external context, *Validation*, which considers "the alignment of data values with respect to relevant external benchmarks". Our analysis addresses *Verification* only. The Kahn framework gives detailed definitions for DQ measures in each category. For each of the proposed definitions, we analyzed which CDS elements would be required to implement a DQ measure. The resulting formalizations are independent of any DQA tool implementation. Therefore, we refer to them as data quality indicators (DQIs).

All implementations for DQA in the 6th Projectathon were based on those DQIs. The R-Tool *firecracker*[4] was used to flatten the data as was required by the analysis scripts. We used the R-Package *dataquieR* [3] to perform our DQA. First, we generated the metadata in the format required by *dataquieR*. Secondly, we wrote an R-Script that utilizes the methods provided by *dataquieR* to perform the actual DQA. This R-Script is called during the data-flattening process, and automatically generates a DQ-Report, which now includes results for each of the initially formalized DQIs.

3. Results

Table 1 shows the formalized DQIs for each of the definitions established by Kahn et al. For each definition, we describe relevant elements from the CDS profiles. We also list which DQIs were evaluated during the 6th Projectathon.

The Kahn *Conformance* category was of particular interest for the CDS, as running a FHIR validator should ensure full conformance with the CDS profiles. But especially for this category, processes in the hospital can result in DQ concerns. In some hospitals, the FHIR *Observation* resource is created as soon as a measurement is requested at the hospital laboratory. The same *Observation* resource should later be updated with the measurement value. If this is not possible, a new *Observation* is created instead. This conveys what had happened in the hospital and will be in accordance with the profile during FHIR validation, but violates *Relational Conformance (b),* which addresses such duplicated resources. This requires further investigation in appropriate provenance concepts specifically for ETL processes and associated data sets at the DICs [6].

A surprising problem occurred in the *Plausibility* category. The CDS captures laboratory measurements in *Observations* occurring during an *Encounter*. While the time at which a laboratory measurement was taken is undisputable, the end date of an *Encounter* is not. Patient discharge, recurring patients, outpatient care and billing necessities influence an *Encounter's* start and end time. As such processes can vary between hospitals, violations of *Temporal Plausibility (b)*, which addresses the correct order of sequences, might be more indicative of process differences, than of errors during the laboratory measurement or data capture for *Observations*.

[4] https://github.com/POLAR-fhiR/fhircrackr

Table 1. Applicable CDS profiles and elements for each definition from the Kahn Framework. Definitions are represented by one letter, according to the table provided by Kahn et al. [5, p. 7-8] (verification context).

Kahn Definition	Description	CDS Profile (CDS Elements)	Projec-tathon
Value Conformance (a)	The value consists only of digits and dot.	Observation (NT-proBNP.value)	yes
Value Conformance (b)	Gender consists only of the allowed categorical values (HL7 Administrative Gender).	Patient (Patient.gender)	yes
Relational Conformance (a)	The references are resolvable in both directions (if any), Conditions are determined by the Encounter.	Condition (Condition.id) Encounter (Encounter.id)	yes
Relational Conformance (b)	Do two Encounters with the same start and end date exist with different IDs?	Encounter (Encounter.startdate, Encounter.enddate) *	yes
Relational Conformance (c)	An element from any older MII Core Data Set profile which had been changed since a previous version.	any CDS Profile with different Elements in a previous version	no
Computational Conformance (a)	The calculation of the age from the date of birth by us is equal to the calculated age in the DIZ.	Patient (Patient.birthdate)	no
Completeness (a)	Gender should not be empty.	Patient (Patient.gender)	yes
Completeness (b)	Patients with multiple encounters should have a certain chronic diagnosis in each medical case.	Condition (Condition.id)* Encounter(Encounter.id)*	no
Uniqueness Plausibility (a)	Each individual Encounter should only be assigned to one unique patient.	Patient (Patient.id) * Encounter (Encounter.id)	yes
Temporal Plausibility (a)	The Encounter's end date should not occur before the Encounter's start date.	Encounter (Encounter.startdate, Encounter.enddate)	yes
Temporal Plausibility (b)	The date of the NT-proBNP measurement is between the corresponding Encounter's start and end date.	Encounter (Encounter.startdate, Encounter.enddate) Observation (NT-proBNP.date)	yes
Temporal Plausibility (c)	Chronic diagnoses of a patient gain diagnostic confidence over time.	Condition (Diagnosis.verificationStatus.code)*	no
Atemporal Plausibility (a)	Hard Limits: NT-proBNP < 10,000 pg/ml	Observation (NT-proBNP.value)	yes
Atemporal Plausibility (b)	Difference in value of the same measurement obtained by different instruments.	Observation (NT-proBNP.value)*	no
Atemporal Plausibility (c)	Different hard limits for male and female patients.	Observation (NT-proBNP.value) Patient (Patient.gender)	yes
Atemporal Plausibility (d)	Difference in values for NT-proBNP measurement of a patient in case of multiple measurements.	Observation (NT-proBNP.value)*	no

* required multiple times

4. Discussion and Conclusions

Our results show that well-formalized DQIs will not only find erroneous data points but also have the potential to highlight issues in the underlying data-capturing processes. While a comprehensive DQA setup can reveal differences in those underlying processes, what is admissible depends on the specific use case.

The use case of the 6th Projectathon allowed us to implement such a comprehensive DQA setup based on our initially formalized DQIs for ten of the sixteen DQ measures defined by Kahn et al. While the limited data required in that use case prevented further implementations, we used the FHIR profiles of the CDS to formalize DQIs for all sixteen definitions. We hope that this benefits the DQA efforts of future use cases built on the CDS. Moreover, our efforts highlight the benefits of formalizing DQIs independently of the requirements from a specific DQA tool. Their natural language descriptions allow the inclusion of domain experts with detailed knowledge of the processes at data capture during a clinical routine, the ETL processes at the DICs, or the CDS FHIR profiles, but with limited programming skills. The independent conceptualization offers more flexibility in the later implementation. Also, due to this tool agnostic nature these DQIs are potentially relevant towards further improving already established DQA processes and pre-existing tools deployed at any point of the clinical data life cycle. Especially if such processes and tools had been designed for FHIR data, as is the case with FHIR CQL, which sees some use outside of the MII already. Lastly, the conceptual model can be updated independently of the DQA tool.

For the MII, if those DQIs can be stored in an interoperable, reusable way, they could be beneficial for initial DQA conducted directly at the DICs. In the future, we hope that such a *FAIRification* of independent DQIs can also help to address the currently omitted *Validation* context. Finally, this will enable gold-standard references to be built on the extensive data stored at DICs.

Acknowledgments: Research was supported by BMBF grants SMITH (01ZZ1803A), POLAR (01ZZ1910A) and DFG grants WI 1605/10-2, NFDI 13/1.

References

[1] Gehring S, Eulenfeld R. German Medical Informatics initiative: Unlocking data for Research and Health Care. Methods of Information in Medicine. 2018;57(S 01). doi:10.3414/me18-13-0001.

[2] Kohane IS, Aronow BJ, Avillach P, et al. What every reader should know about studies using electronic health record data but may be afraid to ask. Journal of Medical Internet Research. 2021;23(3). doi:10.2196/22219.

[3] Schmidt CO, Struckmann S, Enzenbach C, et al. Facilitating Harmonized Data Quality Assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Medical Research Methodology. 2021;21(1). doi:10.1186/s12874-021-01252-7.

[4] Kapsner LA, Mang JM, Mate S, et al. Linking a consortium-wide data quality assessment tool with the MIRACUM Metadata Repository. Applied Clinical Informatics. 2021;12(04):826–35. doi:10.1055/s-0041-1733847.

[5] Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and framework for the secondary use of electronic health record data. eGEMs. 2016;4(1):18. doi:10.13063/2327-9214.1244.

[6] Gierend K, Krüger F, Waltemath D, et al. Approaches and criteria for provenance in biomedical data sets and workflows: Protocol for a scoping review. JMIR Research Protocols. 2021;10(11). doi:10.2196/31750.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230118

Automating Structured Results Communication to Expedite Imaging-Directed Care in Spine Oncology

Alvand HASSANKHANI [a], Raghav R. MATTAY [a], Colbey W FREEMAN [a], Charles E. KAHN Jr. [a,1]
[a] *Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA*
ORCiD ID: A. Hassankhani https://orcid.org/0000-0003-1292-5265
C.E. Kahn https://orcid.org/0000-0002-6654-7434

Abstract. Cancers frequently metastasize to the spine, where they can cause severe morbidity, including pain, vertebral collapse, and paralysis. Accurate assessment and timely communication of actionable imaging findings are critical. We developed a scoring mechanism to capture the key imaging features of examinations performed to detect and characterize spinal metastases in patients with cancer. An automated system was developed to relay those findings to the institution's spine oncology team to expedite treatment. This report describes the scoring scheme, the automated results communication platform, and initial clinical experience with the system. The scoring system and communication platform enable prompt, imaging-directed care of patients with spinal metastases.

Keywords. Oncology, Radiology, Results communication, Structured reporting, Common Data Elements

1. Introduction

Metastasis to the spine is a frequent complication of many forms of cancer: it is estimated that more than 30% of cancer patients will develop spinal metastasis [1]. Spinal metastases can cause severe pain due to nerve compression or vertebral collapse, and can result in paralysis due to compression of the spinal cord. The proper management of patients requires a timely multi-specialty approach [2,3]. Therefore, it is important that spinal metastases be communicated promptly to the physicians who can manage the disease. Even in a tertiary care medical center, where the medical staff order, perform, and interpret spine imaging examinations promptly, there can be unintended delays in communicating actionable findings to the neurosurgeons, orthopaedic surgeons, and radiation oncologists who are prepared to treat the disease.

The Spine Instability Neoplastic Score (SINS) and the Epidural Spinal Cord Compression (ESCC) scale provide imaging and clinical characteristics that are critical in treatment decision-making in patients with spinal metastatic disease, and have been validated in clinical practice [4-7]. We created a scoring scheme based on magnetic resonance imaging (MRI) features of SINS and ESCC for spinal column stability and

[1] Corresponding Author: Charles E. KAHN Jr, E-mail: ckahn@upenn.edu.

cord compression; we omitted clinical features, such as mechanical pain, and imaging features that would be difficult to determine on MRI, such as the lytic or blastic nature of the metastases. We also developed and implemented a novel system to expedite highly detailed information about spinal metastatic disease to our institution's Spine Oncology treatment team.

2. Methods

2.1. Spine Oncology Imaging Score

In collaboration with our institution's spine surgeons, neurologists, and radiation oncologists, the radiology department formulated the Spine Oncology Imaging Score (SOIS) to highlight discrete MRI features from the SINS and ESCC scoring methods. To make SOIS readily available for use by the radiology department, we encoded it as a text "macro" within the institution's radiology reporting system (Nuance PowerScribe 360, Microsoft, Redmond, WA). By invoking the "Spine Oncology" macro, a radiologist can incorporate predefined text into the radiology report of a spine MRI examination. The text macro includes pick-list fields that allow selection of discrete common data elements to capture the SOIS components (Table 1).

Table 1. Components of the Spine Oncology Imaging Score.

Component	Values
Location	Rigid spine (S2-S5): 0 points Semi-rigid spine (T3-T10): 1 point Mobile spine (C3-C6, L2-L4): 2 points Junctional spine (Occiput-C2, C7-T2, T11-L1, L5-S1): 3 points
Alignment	Normal: 0 points Deformity (kyphosis/scoliosis): 2 points Subluxation: 4 points
Collapse	None: 0 points No collapse with > 50% vertebral body involvement: 1 point < 50%: 2 points > 50%: 3 points
Posterior elements	None: 0 points Unilateral: 1 point Bilateral: 3 points
Compression grade	Grade 0 (osseous disease only) Grade 1a (epidural involvement without thecal sac deformity) Grade 1b (thecal sac deformity without cord contact) Grade 1c (thecal sac deformity with cord contact) Grade 2 (cord compression with preservation of some CSF) Grade 3 (cord compression with complete effacement of CSF) Grade 4 (intramedullary/cord or drop metastasis) Grade 5 (intradural and extradural)

2.2. Notification system

To promote rapid dissemination of actionable imaging findings to the appropriate care team, we instituted an automated notification system that is invoked hourly. Scripts in Perl version 5.24.1 on a Linux x86-64 platform incorporated three Perl modules:

DBI::data to connect to the main database, MIME::Lite to send notification emails, and HTML::Template to compose emails in HTML. Data for the emailed notifications was compiled using a stored procedure written in the Structured Query Language (SQL). The stored procedure first searches the Nuance PowerScribe 360 (Microsoft, Redmond, WA) production database for any studies where the Spine Oncology macro was used. The script also searches the report text of all spine MRI exams for the string "[SOIS21]." For reports that meet the search criteria, the alert messages include the patient's name and medical record number, current location, and full text of the imaging report (Figure 1). The alerts are transmitted as secure electronic mail messages within the institutional environment protected by a multi-layer firewall; as all treatment specialists belong to the same healthcare organization, no external communication is made.

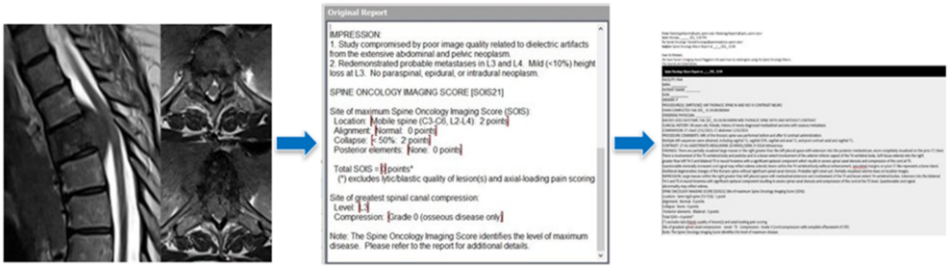


Figure 1. Sample workflow. Sagittal and axial MR images of the thoracic spine demonstrate a new T5 vertebral body metastasis with epidural spinal cord compression. The radiologist incorporates the SOIS macro into the radiology report. Less than an hour after the report was rendered, a secure electronic message is transmitted to the spine oncology specialists.

```
SPINE ONCOLOGY IMAGING SCORE [SOIS21]
  Site of maximum Spine Oncology Imaging Score (SOIS): T11
  Location: Junctional spine (Occiput-C2, C7-T2, T11-L1, L5-S1): 3
        points
  Alignment: Normal: 0 points
  Collapse: > 50%: 3 points
  Posterior elements: Bilateral: 3 points

Total SOIS = 9 points*
(*) excludes lytic/blastic quality of lesion(s) and axial-loading pain
        scoring

Site of greatest spinal canal compression: Level: T11
Compression: Grade 3 (cord compression with complete effacement of CSF)

Note: The Spine Oncology Imaging Score identifies the level of maximum
disease. Please refer to the report for additional details.
```

Figure 2. Portion of an example report showing use of the Spine Oncology text macro.

3. Clinical Implementation

From April 2021 to January 2023, the SOIS macro was incorporated into 1017 MR imaging exams of the cervical, thoracic, or lumbar spine in 431 patients; 491 of the exams (48%) were in female patients. 161 exams were referred from the emergency department,

365 on inpatients, and 490 on outpatients (Figure 3). Distribution by age of reports incorporating the Spine Oncology macro is shown in Figure 4.

Radiologists incorporated the Spine Oncology text macro into all relevant radiology reports, and the automated system functioned properly to send electronic notifications to the spine oncology multidisciplinary team within 1 hour of either the preliminary or final radiology report.

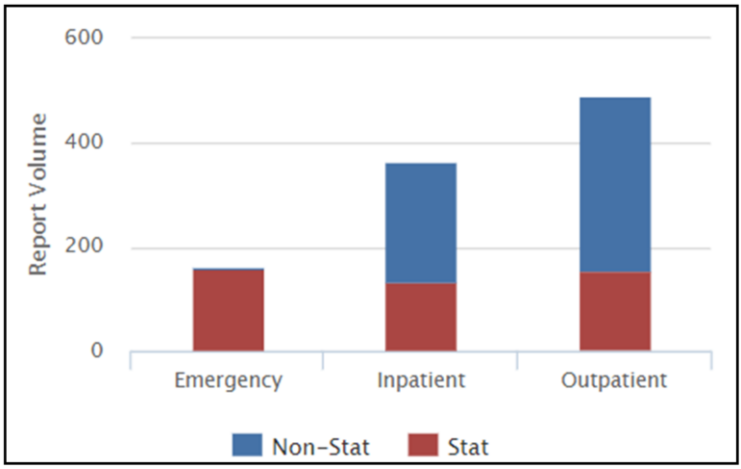


Figure 3. Distribution of exams by patient-care setting for reports that included the Spine Oncology text macro.

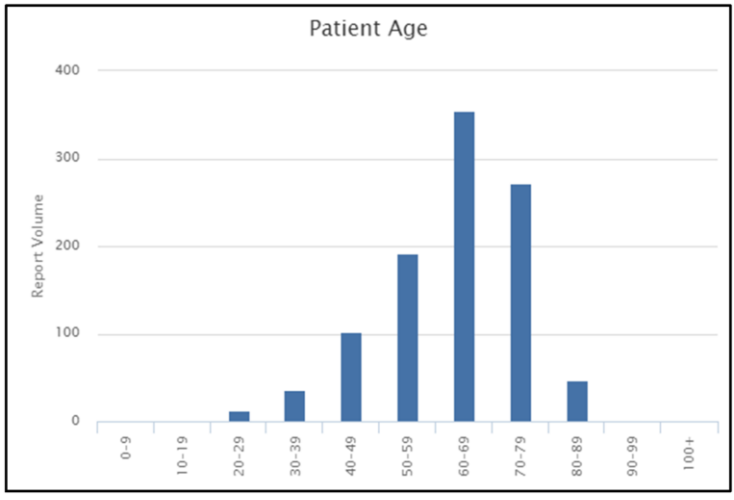


Figure 4. Distribution of exams with Spine Oncology reporting by patient age.

4. Ongoing Work

The Spine Oncology text macro includes a "tracker code"—an arbitrary sequence of letters and digits within square brackets, here "[SOIS21]"—which enables on-demand retrieval of relevant reports. A radiology report text-search system (Nuance mPower, Microsoft, Redmond, WA) retrieves reports that include the tracker code (and hence, the

Spine Oncology macro) for any specified time period. The data available for retrieval are updated every 24 hours. For exams flagged with the Spine Oncology macro, the management plan and outcome of each patient are reviewed and discussed at a weekly multidisciplinary conference. The use of the SOIS macro for more than 1,000 examinations in our multi-hospital tertiary academic institution is a testament to its utility across the several specialties that provide care to patients with spinal metastasis.

SOIS data elements, as part of the ESCC scale, have been incorporated into a library of Common Data Elements (CDEs) for radiology (https://radelement.org/set/RDES21), developed by the Radiological Society of North America (RSNA) and the American College of Radiology (ACR) [8]. CDEs define the attributes and allowable values of a unit of information, so that data can be collected and stored uniformly across institutions and studies. The RadElement.org data dictionary defines radiology CDEs is to make these data interoperable for a variety of applications, including clinical reports, computer-aided reporting systems, computer vision applications, clinical research report forms, and radiology case collections.

5. Conclusion

The Spine Oncology Imaging Score system, together with the electronic referral pipeline, has been technically successful and practical in automating referrals. By generating an electronic notification shortly after the radiology report is generated, the time interval between imaging and consultation is significantly reduced, which has enabled expedited and seamless patient care of these patients with time-critical imaging findings. The simple, image-based oncology score not only adds value but directs care to cancer patients with an unstable spine.

References

[1] Van den Brande R, Cornips EM, Peeters M, Ost P, Billiet C, Van de Kelft E. Epidemiology of spinal metastases, metastatic epidural spinal cord compression and pathologic vertebral compression fractures in patients with solid tumors: A systematic review. *J Bone Oncol* 2022;35:100446. doi: 10.1016/j.jbo.2022.100446.

[2] Nowak H, Szwacka DM, Pater M, et al. Holistic approach to the diagnosis and treatment of patients with tumor metastases to the spine. *Cancers* 2022;14(14). doi: 10.3390/cancers14143480.

[3] Barzilai O, Fisher CG, Bilsky MH. State of the art treatment of spinal metastatic disease. *Neurosurgery* 2018;82(6):757-69. doi: 10.1093/neuros/nyx567.

[4] Fisher CG, DiPaola CP, Ryken TC, et al. A novel classification system for spinal instability in neoplastic disease: an evidence-based approach and expert consensus from the Spine Oncology Study Group. *Spine* 2010;35(22):E1221-9. doi: 10.1097/BRS.0b013e3181e16ae2.

[5] Quraishi NA, Arealis G, Salem KM, Purushothamdas S, Edwards KL, Boszczyk BM. The surgical management of metastatic spinal tumors based on an Epidural Spinal Cord Compression (ESCC) scale. *Spine J* 2015;15(8):1738-43. doi: 10.1016/j.spinee.2015.03.040.

[6] Fourney DR, Frangou EM, Ryken TC, et al. Spinal instability neoplastic score: an analysis of reliability and validity from the spine oncology study group. *J Clin Oncol* 2011;29(22):3072-7. doi: 10.1200/jco.2010.34.3897.

[7] Bilsky MH, Laufer I, Fourney DR, et al. Reliability analysis of the epidural spinal cord compression scale. *J Neurosurg Spine* 2010;13(3):324-8. doi: 10.3171/2010.3.spine09459.

[8] Flanders AE, Jordan JE. The ASNR-ACR-RSNA Common Data Elements Project: What will it do for the house of neuroradiology? *American Journal of Neuroradiology* 2019;40(1):14-18. doi: 10.3174/ajnr.A5780.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230119

Greek Hospital Data Mining and Analysis

Maria CHINTIROGLOU[a,1], Haralampos KARANIKAS[a] and Sotiris TASOULIS [a]
[a] *Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece*

Abstract. Monitoring the performance of hospitals is a crucial issue related both with the quality of healthcare services and with country's economy. An easy and trustful way of evaluating health systems is through key performance indicators (KPIs). Such indicators are widely used for the identification of gaps in the quality or efficiency of the services provided. The main aim of this study is the analysis of the financial and operational indicators at hospitals in the 3rd and 5th Healthcare Regions of Greece. In addition, through cluster analysis and data visualization we attempt to uncover hidden patterns that may lie within our data. The results of the study support the need for re-evaluation of the assessment methodology of Greek hospitals to identify the weaknesses in the system, while evidently unsupervised learning exposes the potential of group-based decision making.

Keywords. Hospital Data, Data Analysis, eHealth, key performance indicators.

1. Introduction

Over the years, various methods have been sought to ensure the proper performance of hospitals and increase their efficiency in Greece. Despite the scientific and technological development, the health systems of many countries still present weaknesses. It is estimated that 40%-80% of a country's total expenses are allocated to hospitals [1]. Therefore, it was deemed necessary to adopt policies for evaluating hospitals' operation.

In Greece, one of the methods for monitoring the operation of hospitals is by certain established KPIs. These indicators can be divided in two main categories: financial and operational [2]. The financial indicators include annual expenditure for raw and auxiliary materials (pharmaceuticals, hygiene supplies, orthopedic equipment, reagents, etc.), consumables (gas, fuel, etc.) and detailed data on salaries, payments, and revenues for every hospital. The hospital activity (operational) data include the number of inpatients and the total hospitalization days, the number of outpatients and the emergency services for every hospital. In this study, data from hospitals in the 3rd and 5th healthcare regions of Greece were used. According to the data of 2014, the corresponding indicators have been produced and then evaluated. In addition, particular indicators have been combined to construct a new data matrix, upon which unsupervised learning have been applied to discover groups of hospital with similar respective behavior. While the results are specific, this method is rather generic and can also be employed in other countries, setups, or data.

[1] Corresponding Author: Maria Chintiroglou, Department of Computer Science and Biomedical Informatics, University of Thessaly, Papasiopoulou 2-4, Postal Code: 35131, Lamia, Greece; E-mail: mhintirog@gmail.com

2. Methods

The provided data, financial data of the hospitals (1), data corresponding to the number of beds (2), data relative to the inpatient number (3), and data concerning the outpatients (4), underwent a pre-screening process, to facilitate the analysis of the indicators and the clustering procedure.

Our basic workflow was divided in two main parts:

- The production and the evaluation of indicators.
- The combination of indicators, and the subsequent cluster analysis.

This study estimated and examined the following indicatively KPIs: average cost per patient, defined as the total expenditure (expenditure for raw and auxiliary materials and consumables etc., excluding payroll) of a hospital, divided by the number of hospitalized patients; average cost per hospital day, defined as the total expenditure (expenditure for raw and auxiliary materials and consumables etc., excluding payroll) of a hospital, divided by the total number of hospital days; average drug cost per patient, defined as the total pharmaceutical expenditure of the hospital, divided by the number of hospitalized patients. Initially, we calculated the total number of patients, days of hospitalization, and of beds, needed to produce the corresponding indicators. Linear regression analysis was performed to select between the number of patients and the number of days of hospitalization as the most suitable for the calculation of each indicator. The most appropriate indicators are being utilized to generate the dataset for the unsupervised analysis. The resulting dataset is constituted by the 7 calculated indicators along with a variable with the total number of beds per hospital. It is important to mention that for the creation of the dataset we only considered general hospitals to focus our analysis on the dominant type of hospitals found at the data at hand. The inclusion of other types could provide misleading results due to the major fundamental differences between hospital types. To uncover hidden patterns that may lie within our data we utilize clustering. Initially, we separate the data into group of hospitals that share similar characteristics and then investigate the critical factors that determine the retrieved groups (clusters). Cluster analysis, combined with visualization of the results, is a very popular and well-established methodology usually used to make sense of large data volumes, finding wide applicability even on recent applications in Bioinformatics and Big Data [3, 4, 5]. Although there are several recent advancements in clustering algorithms [6, 7], for the task at hand we utilize the traditional “k-means” due to its simplicity and control over its parameters. An adequate amount of data would allow the utilization of more sophisticated techniques that promote explainability even further. To this end, the available dataset’s manageable scale allows multiple algorithmic executions that can lead to relative optimal results, without the need for advanced variations intended for complexity reduction and appropriate initialization.

3. Results

In this section, the charts of the most important and most discussed hospital indicators, namely average total cost per patient and pharmaceutical cost per patient (Fig.1 and Fig.2, respectively), are indicatively presented.

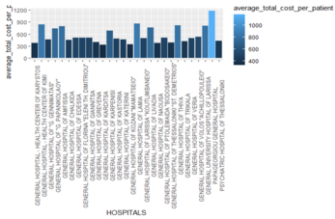


Figure 1. Average total cost per patient.

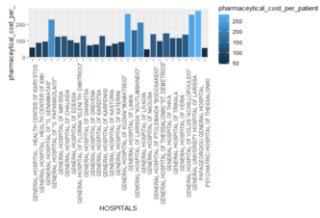


Figure 2. Pharmaceutical cost per patient.

From Figure 1 it is clear that Papageorgiou Hospital has the highest average total cost per patient and, on the opposite side, Thiva Hospital has the lowest. In general, most of the hospitals have similar average inpatient costs except for some that deviate highly. It is noteworthy that Kymi General Hospital has a fairly high average cost, due to its restricted facilities and more specifically based on its low number of beds. This could be possibly attributed to the aeromedical evacuation that may be required, as Kymi is a relatively remote place in Greece. From Figure 2 it can be observed that the largest number of hospitals show similar average costs for drugs with only five hospitals differing in a larger value. Papageorgiou General Hospital shows the highest pharmaceutical cost per patient, and opposingly, Naoussa General Hospital shows the lowest. Pharmaceutical costs are certainly influenced by many factors such as the type of cases that each hospital deals with, which may explain the large variations between hospitals. Moreover, the average length of stay is around 4 days in all hospitals apart from the Psychiatric Hospital. This is reasonable because psychiatric cases probably need a greater length of stay and care [8].

The occupancy of the beds, as it was observed, is greatly increased in Karystos and Kymis Hospitals. After all, they are the hospitals with the lowest number of beds. In the hospitals with the largest number of beds, such as Papageorgiou Hospital and Papanikolaou Hospital, the bed occupancy is much lower. Additionally, very low bed occupancy is observed in the University Hospital and the Psychiatric Hospital.

To this end we employ the dataset described in “Methods” Section. Initially, we perform range normalization across all variables bounding their values in [0,1], while to determine the number of clusters parameter we utilize the elbow method. In figure 3 we see a line chart of the SSE for each value of k. Based on the results and on the fact that we need to retain a relatively low number of clusters for straightforward interpret-ability we set k = 4. To visually investigate the uncovered data structure we perform Principal Components Analysis (PCA) to project the data onto the first two principal components. As shown, in Figure 3 we can clearly distinguish identified clusters.

Figure 3. Principal Components Analysis (PCA).

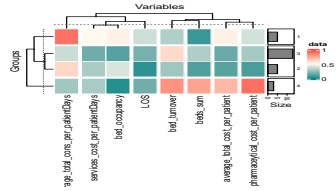


Figure 4. Average values for each normalized variable per cluster.

To identify the aspects of each retrieved cluster we present the corresponding average values of each variable in Figure 4, allowing us to discriminate major differences along clusters. To enhance the visualization both retrieved clusters and variables are grouped through agglomerative clustering to re-positioning them according to their in-between similarities. The size of each cluster is also reported through a bar plot.

Finally, we attempt to relate the clustering result with the hospital geolocation in order identify the relation of the extracted hospital factors with their geographical location. Figure 5 illustrate the location of all hospitals colored according to the respective cluster label of Figure 3.

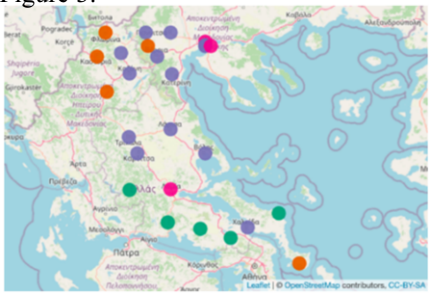


Figure 5. Hospitals Geographic Location.

From the above clustering we could proceed to draw some conclusions. First of all, we conclude that hospitals belonging to nearby cities have common characteristics in their mode of operation based on their indicators. This could be either because of the life of the citizens in each area or because of following a common strategy in the operation of the nearby hospitals. We also observe that the number of beds has a significant impact on the performance of hospitals. We observe that hospitals with the largest number of beds have lower costs for services and for materials and consumables and higher costs for drugs while hospitals with the smallest number of beds have higher costs in materials and consumables and lower costs for drugs. This could be due to the fact that medical cases requiring more serious care such as more serious operations or admissions to ICU etc. are transferred to the larger and more centralized hospitals in the country. This affects the larger hospitals that must manage more serious cases and thus increasing their overall costs per patient and their overall costs for drugs. Smaller hospitals on the other hand are burdened with costs for materials and consumables such as costs for patient transport etc. [9]. Furthermore, it seems that bed occupancy, a very important indicator for the performance of a hospital, is affected by the size of the number of beds, as hospitals with the largest number of beds do not have a problem with occupancy while hospitals with the lowest number of beds seem to have an issue with their bed occupancy which may also support our previous conclusion about the transfer of patients from smaller to larger hospitals also due to their bed occupancy.

4. Discussion and Conclusions

Hospitals have to face many cases of patients daily, such as outpatients, emergency cases, etc., with high monitoring needs. Due to the economic crisis, the workload requires faster and more trustworthy data processing. Thus, a way to monitor hospitals' data in a faster and more reliable fashion, is with the use of indicators. Furthermore, extracting knowledge from this valuable data is a major piece of research. The use of clustering

algorithms can contribute to optimal and faster knowledge discovery. The need for monitoring and evaluation of health units is, therefore, urgent. Indicators are a reliable way to achieve this. From the indicators calculated above, we can conclude that in general the larger hospitals face the most workload and have the highest costs apart from some exceptions such as Kymi General Hospital which is a small hospital but has high costs for services, materials, and consumables. At the same time, it can be concluded that the average length of stay in almost all hospitals is not very long, with an average of 3-5 days, except for Psychiatric hospitals, where patients stay on average 20 days.

Furthermore, it appears that hospitals that belongs in nearby geographical areas exhibit similar characteristics in their operation. Hospitals that have similar number of beds also show similar performance. Finally, we could conclude that smaller and relatively remote hospitals face more difficulties in their performance than hospitals located in larger and more central cities of the country. Finally, all the above observations and conclusions allow us to gain an insight into the quality of care in each hospital. The bed occupancy rate, for example, which was mentioned above, is a very important indicator not only for the functioning of a hospital, but also for the quality of care it provides. Hospitals with a high bed occupancy rate, above average and above permissible limits, cannot provide the same quality level of care to their patients as hospitals that do not face such a high workload.

References

[1] Rahimi H, Khammar-nia M, Kavosi Z, Eslahi M. Indicators of hospital performance evaluation: a systematic review. International Journal of Hospital Research. 2014 Dec;3(4):199-208.

[2] Christodoulakis A, Karanikas H, Billiris A, Thireos E, Pelekis N. “Big data” in health care Assessment of the performance of Greek NHS hospitals using key performance and clinical workload indicators. Journal: ARCHIVES OF HELLENIC MEDICINE. 2016;33(4):489-497.

[3] Ianni M, Masciari E, Mazzeo GM, Mezzanzanica M, Zaniolo C. Fast and effective big data exploration by clustering.Future Generation Computer Systems.2020;102:84-94.doi: https://doi.org/https://doi.org/10.1016/j.future.2019.07.077

[4] Kern M, Lex A, Gehlenborg N, Johnson CR. Interactive visual exploration and refinement of cluster assignments. BMC Bioinformatics. 2017 Sep;18(1). doi: https://doi.org/10.1186/s12859-017-1813-7

[5] Ostaszewski M, Kieffer E, Danoy G, Schneider R, Bouvry P. Clustering approaches for visual knowledge exploration in molecular interaction networks. BMC Bioinformatics. 2018 Aug;19(1).doi: https://doi.org/10.1186/s12859-018-2314-z.

[6] Sieranoja S, Fra¨nti P. Fast and general density peaks clustering. Pattern Recognition Letters. 2019 Dec; 128:551-558. doi: https://doi.org/10.1016/j.patrec.2019.10.019.

[7] Tasoulis S, Pavlidis NG, Roos T. Nonlinear dimensionality reduction for clustering. Pattern Recognition. 2020 Nov;107:107508. doi: https://doi.org/10.1016/j.patcog.2020.107508.

[8] Bressi SK, Marcus SC, Solomon PL. The impact of psychiatric comorbidity on general hospital length of stay. Psychiatric Quarterly. 2006 Sep;77(3):203-209.doi: https://doi.org/10.1007/s11126-006-9007-x

[9] Kontodimopoulos N, Nanos P, Niakas D. Balancing efficiency of health services and equity of access in remote areas in Greece. Health policy. 2006;76(1):49-57.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230120

Challenges with Medication Management and the National Medication List in Sweden: An Interview Study from a Human, Organizational, and Technology Perspective

Tora HAMMAR [a,1], Mikael HOFFMANN [b] and Lina NILSSON [a]
[a] *eHealth institute, Department of Medicine and Optometry, Linnaeus University*
[b] *The NEPI Foundation, Stockholm, Sweden*
ORCiD ID: Tora Hammar https://orcid.org/0000-0003-1549-2469 , Mikael Hoffmann https://orcid.org/0000-0001-9420-3316 , Lina Nilsson https://orcid.org/0000-0003-2074-3584

Abstract. Sweden is in the process of implementing the National Medication List (NLL). The aim of this study was to explore the challenges with the medication management process, as well as expectation for NLL, from a human, organizational, and technology perspective. This study included interviews with prescribers, nurses, pharmacists, patients, and their relatives and was conducted during March to June 2020, before the implementation of NLL. Challenges were (1) feeling lost with several different medication lists, (2) spending time searching for information, (3) being frustrated at parallel information systems, (4) patients being the carriers of information, and (5) the feeling of being responsible in an indistinct process. The expectations for NLL in Sweden were high, but there were several fears.

Keywords. Medication, shared information, sociotechnical perspective, interview

1. Introduction

The lack of a shared and up-to-date medication list can create uncertainty among patients, increases the risk of inappropriate combinations of medicines, as well as creating unnecessary extra work [1]. Sweden is in the process of introducing the National Medication List (NLL), based on a law that came into force on May 1, 2021 [2, 3]. The implementation is done in steps and the preliminary date for integration in all EHRs is December 2025. NLL is part of a complex system of people, technology and organizations where effects can influence and be influenced by many different aspects [4]. The goal for NLL is to give healthcare, pharmacies and patients access to the same information about prescribed and dispensed medications, i.e., expressed from a technological perspective (T). The aim of this study was to explore the challenges with

[1] Corresponding Author: Tora Hammar, eHealth Institute, Department of medicine and optometry, Linnaeus University, 391 82 Kalmar, Sweden, E-mail: tora.hammar@lnu.se.

the medication management process in Sweden from a human, organizational, and technology perspective. Also, this study aimed at exploring expectations of NLL in Sweden.

2. Methods

This interview study with health professionals, patients and relatives was performed during March to June 2020, before the 1st step of implementation of NLL. In total, 33 informants (23 women and 10 men) participated in this study (Table 1). The individual interviews were mainly performed via phone, but a few were done physically or via video, and lasted between 27 and 97 minutes (mean 41 minutes). Strategic selection was used to include many different perspectives. Informants lived in different parts of Sweden and worked in different sectors and handled medication either professionally or in their personal life. Patients had three or more medications, and relatives helped a family member with medications (ages from 33 to 69). A deductive data analysis was conducted where data from transcribed interviews were coded to the pre-defined categories Human (H), Organization (O), Technology (T) [5]. Thereafter subcategories were identified and classified as belonging to any of the HOT-categories or the subsection between them using an inductive method. This study was approved by the Swedish Ethical Review Authority (Dnr 2019-06553).

Table 1. Participant demographics (n=33).

Informants	Number (n=33)
Patient (three or more medications)	10
Relatives (helping close family with medication)	3
Pharmacist (community pharmacy)	8
Clinical pharmacist (health care)	2
Physician (health care)	7
Registered nurse (health care)	3

3. Results

In the analysis, six subcategories were identified based on the categories Human (H), Organization (O), and Technology (T) and the subsection between them (Figure 1). Three subcategories belonged to the middle section where all three categories overlap, and three subcategories belonged to the intersection between two of the categories.

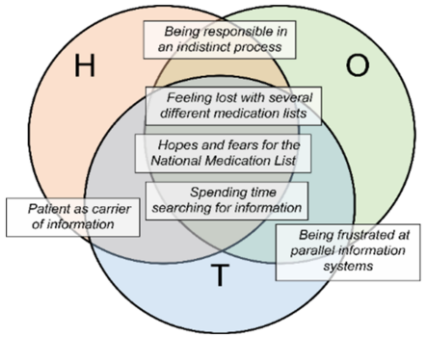


Figure 1. Categories Human (H), Organization (O), and Technology (T) and six identified subcategories.

3.1. Patient as carrier of information

Almost all patients and close family described that they felt they had to memorize and be the carrier of the information about medication they were prescribed. They reported several different tricks and tools, both digital and manual to remember what medications and when to take medications. Tricks and tools were seen as ways of feeling safe and in control of their medication treatment. Health care professionals also experiences that they had to rely on the patient (or their relatives) as carrier of information about medication treatment.

[…] you create your own system to find experienced control over your own medication. (Patient 9)

You try to connect all sources you may get and then you check it with the patient. (Physician 5)

3.2. Being responsible in an indistinct process

The medication management process was experienced by several informants to include gaps in information flow. Physicians, pharmacists, and nurses experienced that the responsibility were divided between them as professions and sometimes also with the patient. Because of incorrect lists or missing information professions stated that they needed to contact each other to double check information. Communication was expressed as challenging since professionals in different parts of the process use different systems and insufficient tools and routines for communication.

'[…] very slow […] somewhere you wish there was a two-way communication between pharmacies and health care where you could be engaged in a dialogue or at least get a direct contact with them [health care], that you could write a question and get the prescription back' (Pharmacist 8)

Patients and relatives experienced that health care and pharmacy took responsibility for the medication. Though, when medications were changed in some way, several patients and relatives expressed worries of the medication management being less reliable.

3.3. Feeling lost with several different medication lists

Almost all informants referred to having different medication lists. Sometimes patients and close family had their own digital or manual medication list. Often, physicians, pharmacists, and nurses referred to two different lists, the medication list in the health care EHR, and the prescription list available at pharmacies (called the National prescription repository, when the study was done).

'The medication list is the current treatment of the patient; the medication you take now and why […] your medication schedule from your physician you might say. The prescription list is a list of your current [valid] prescriptions.' (Pharmacist 4)

Although there were different lists in the medication management process, several informants stated that they trusted their list. At the same time, it was described as very common with errors and discrepancies between the lists. Some informants also used a parallel information system for patients with multi-dose drug dispensing (MDDD).

3.4. Spending time searching for information

Due to gaps in the medication management process, informants experienced the need to search and double-check information with different professions, different organizations, and with patient and sometimes close family. Several described this as a time-consuming

detective work. Searching for information included i.e., searching for interactions, changes prescriptions and dosage. Also, informants expressed that generic exchange (changing brand names at the pharmacy) contributed to increased challenges and risks for patients. Informants referred to different decision-support systems including their medication experience, digital tools, and verbal consultations.

'Well, we use the well-known Google, or 1177 and other reliable sources, you find good information there. Well, it is easy to use if you are wondering about anything, […] It is when we want to make sure, well that we take [the medication] the right way, the dosage […] otherwise you have health care, you can call health care and ask them, and I think that works fine.' (Close family 1)

3.5. Being frustrated at parallel information systems

In the medication management process, informant identified different information systems that were not able to share information. At the same time, informants were not able to see all information in one system and sometimes they did not have access to all information systems required in the process. Some informants stated that medications from different healthcare organization in the medication management process were not included in the same medication list. Different systems were expressed as confusing for patients, close family, and professions and may be seen as a source of errors of different kind.

'If the patient has been hospitalized, they use the ordinary EHR. And if the patient has MDDD that list [in another system] must be adjusted according to medication orders in the EHR. Sometimes they have not added or removed medications properly, so you have to check and compare. Many times, you have to contact the hospital to ask them to adjust.' (Registered nurse 3)

3.6. Hopes and fears for the National Medication List (NLL)

When this study was conducted, several of the informants had very little knowledge about NLL. Almost none of the patients and relatives knew about NLL before the interviews. Informants stated several hopes and opportunities about NLL, i.e., fewer errors in the lists, increased patient safety, and improved communication. At the same time, fears related to the implementation were expressed. Some informants were concerned about NLL being based on prescriptions rather that medication orders, the new required handling of consent and blocking of information, and the difficulties of integrating NLL. Other expressed concerns were related to the fear that NLL may cause confusion in the starting phase of implementation, and that NLL itself will not solve issues in the process. If a list is not updated and if information and communication does not flow appropriately in the process, the challenges will remain.

The handling of consent [with NLL] is a major issue. I think patients, prescribers and pharmacists will have a really hard time understanding where and when information is locked and when it is not. It is two different laws that collide (Clinical pharmacist 2)

'above all to strive for the patient to understand:' what kind of medication do I take and why do I take them and how do the work for me?' I think that is an important part to, to get the knowledge and understanding for patients to take their prescribed medications.' (Registered Nurse 1)

4. Discussion and Conclusion

Challenges with the medication management process are related to having several different medication lists with errors, working in parallel systems, spending time

searching for information and relying on patients as carriers of information although they often lack appropriate support for this. Some of the challenges are in line previous research [1, 3, 6]. The expectations for NLL in Sweden were high, but there were fears related to the implementation. The expectations for NLL were somewhat unrealistic and not in line with the solution currently being implemented in Sweden. Concerns were related to NLL being based on prescriptions rather than medication orders (i.e., decisions about medication treatment). The medication list in modern EHRs is a compilation of connected sequences of medication orders while the e-prescription is a separate one-way communication. This discrepancy complicates integrating the list of prescriptions in NLL with the EHR in healthcare in a way that secures that the information is up to date [6-8]. Also, no one is responsible for a compiled list of prescriptions from different prescribers who haven't accessed the information. In addition, several actions by prescribers such as a change of dose or ending a treatment do not today result in corresponding changes in the list of prescriptions. These are some reasons why NLL in its current state, cannot be recommended as the medication list that patients should use. This study highlights challenges with medication management and NLL from the perspective of different actors, thus contributing with a broader understanding of the complexity. Internationally there are some results that indicate that a shared medication list may contribute to medication lists being more correct, but more research is needed to understand the effects [7, 9]. The challenges identified in this study are all connected to a combination of human, organizational and technology aspects. Most of the challenges described in this study will probably remain until NLL is integrated with health care EHR in a way so that prescribers are assisted in taking responsibility for the completeness and correctness of the information in NLL without too many extra tasks introduced.

References

[1] Manskow US, Kristiansen TT. Challenges faced by health professionals in obtaining correct medication information in the absence of a shared digital medication list. Pharmacy. 2021 Feb 22;9(1):46.DOI: 10.3390/pharmacy9010046.

[2] E-hälsomyndigheten. Nationella läkemedelslistan. 2021; Available from: https://www.ehalsomyndigheten.se/tjanster/yrkesverksam/nationella-lakemedelslistan/.

[3] Hammar T, Mzil L, Eiermann B. Discrepancies in patients' medication lists from pharmacies in Sweden: an interview study before the implementation of the Swedish National Medication List. International Journal of Clinical Pharmacy. 2022 Oct 28:1-9. DOI: 10.1007/s11096-022-01480-x.

[4] Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). International journal of medical informatics. 2008 Jun 1;77(6):386-98..DOI: https://doi.org/10.1016/j.ijmedinf.2007.08.011.

[5] Elo S, Kyngäs H. The qualitative content analysis process. Journal of advanced nursing. 2008 Apr;62(1):107-15. DOI: 10.1111/j.1365-2648.2007.04569.x.

[6] Bülow C, Flagstad Bech C, Ullitz Faerch K, Trærup Andersen J, Byg Armandi H, Treldal C. Discrepancies between the medication list in electronic prescribing systems and patients' actual use of medicines. The Senior care pharmacist. 2019 May 1;34(5):317-24..

[7] Josendal AV, Bergmo TS, Granas AG. Implementation of a shared medication list in primary care–a controlled pre-post study of medication discrepancies. BMC Health Services Research. 2021 Dec 13;21(1):1335.DOI: 10.1186/s12913-021-07346-8.

[8] Sääskilahti M, Ojanen A, Ahonen R, Timonen J. Benefits, Problems, and Potential Improvements in a Nationwide Patient Portal: Cross-sectional Survey of Pharmacy Customers' Experiences. Journal of medical Internet research. 2021 Nov 3;23(11):e31483.DOI: 10.2196/31483.

[9] Waldron C, Cahill J, Cromie S, Delaney T, Kennelly SP, Pevnick JM, Grimes T. Personal Electronic Records of Medications (PERMs) for medication reconciliation at care transitions: a rapid realist review. BMC Medical Informatics and Decision Making. 2021 Dec;21(1):1-7.DOI: 10.1186/s12911-021-01659-8.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230121

Local Data Quality Assessments on EHR-Based Real-World Data for Rare Diseases

Kais TAHAR [a,1], Raphael VERBUECHELN [b], Tamara MARTIN [c], Holm GRAESSNER [c] and Dagmar KREFTING [a]
[a] *Institute of Medical Informatics, University Medical Center Göttingen, Germany*
[b] *Medical Data Integration Center, University Hospital Tübingen, Germany*
[c] *Centre for Rare Diseases, University Hospital Tübingen, Germany*

Abstract. The project "Collaboration on Rare Diseases" CORD-MI connects various university hospitals in Germany to collect sufficient harmonized electronic health record (EHR) data for supporting clinical research in the field of rare diseases (RDs). However, the integration and transformation of heterogeneous data into an interoperable standard through Extract-Transform-Load (ETL) processes is a complex task that may influence the data quality (DQ). Local DQ assessments and control processes are needed to ensure and improve the quality of RD data. We therefore aim to investigate the impact of ETL processes on the quality of transformed RD data. Seven DQ indicators for three independent DQ dimensions were evaluated. The resulting reports show the correctness of calculated DQ metrics and detected DQ issues. Our study provides the first comparison results between the DQ of RD data before and after ETL processes. We found that ETL processes are challenging tasks that influence the quality of RD data. We have demonstrated that our methodology is useful and capable of evaluating the quality of real-world data stored in different formats and structures. Our methodology can therefore be used to improve the quality of RD documentation and to support clinical research.

Keywords. Data quality, rare disease, healthcare standards, ETL, HL7 FHIR

1. Introduction

The research project "Collaboration on Rare Diseases" (CORD-MI) [1] of the German Medical Informatics Initiative (MII) [2] connects multiple university hospitals to support clinical research with electronic health record (EHR) data in the field of rare diseases (RD). In Europe, RD are defined as diseases that affect less than 5 in 10,000 people [3], therefore multi-site data sharing is required to reach a sufficient number of cases. However, the required integration and transformation of heterogeneous data sources through Extract-Transform-Load (ETL) processes into an interoperable format is a complex and challenging task [4]. Such ETL processes raise concerns about data quality (DQ) issues such as completeness, implausibility and semantic integrity of the final data sets [5]. To ensure an appropriate evidence level of scientific outcomes derived from these data, sufficient DQ is necessary [6]. The source of a potential DQ issue, i.e. the primary documentation or the ETL processes, can be determined by evaluating the DQ before and after ETL processes, as presented in this manuscript.

1 Corresponding Author: Kais Tahar, Georg-August-University, University Medical Center Göttingen, Institute of Medical Informatics, 37075 Gottingen, Germany; E-mail: kais.tahar@med.uni-goettingen.de.

Various data quality frameworks have been proposed in the literature [5]–[8]. However, useful DQ assessments on RD data usually depend on specific user and domain requirements [6]. The aim of the present work is to evaluate the quality of RD data before and after ETL processes using a methodology that takes user and domain specific requirements into consideration as proposed in [6]. In this context, we investigate how the quality of real-world data on RDs can be assessed automatically, which methods and tools can be used to compare the quality of real-world data stored in different data formats and structures, which impacts have ETL processes on the quality of RD data, and report the learned lessons.

2. Methods

We describe the employed data sets and DQ metrics before we present the implemented methods in more detail. Since the terms in related works about DQ are often ambiguous, we use in this paper the terminology presented in [6].

2.1. Data sets and data sources

The base population used for this study encompasses all inpatient cases in 2020 of a university hospital, stored in the patient administration system (PAS). PAS is the central subsystem in the hospital information systems (HIS) responsible for patient admission, patient discharge and medical billing. In effect, all clinical subsystems send the captured patient data to the PAS - in our case the SAP IS-H system [9] is used. From the base population, all cases that were coded with an ICD-10-GM code [10] covered in a reference list are included into the study. This reference list is also available on GitHub repository [11]. It includes 143 diagnoses that can be used for coding RDs. The extracted data items are specified in the MII core data set (MII-CDS) [12]. The MII-CDS defines the semantics of required data items and provides the basis for enabling standardized data exchange as well as harmonized DQ assessments across the CORD-MI network [6].

The selected data sets for this study are stored in Fast Healthcare Interoperability Resources (FHIR) [13] and comma-separated values (CSV) formats. Both data sets capture information about the basic modules of the MII-CDS namely Person, Treatment Case, and Diagnosis [12]. Exemplary data sets in CSV and FHIR are provided in [6, 11]. The first data set (SAP data) used for this study was exported directly from SAP IS-H in CSV format. The second data set (FHIR data) was created using an ETL pipeline that extracts clinical data from SAP IS-H according to the German §21 Hospital Remuneration Act (KHEntgG) [14] and transforms these data into FHIR standard. The resulting FHIR resources follow the MII-CDS as specified in the FHIR implementation guide of CORD-MI [15]. These standardized data are stored in a central FHIR server that provides multiple types of FHIR resources such as Patient, Encounter, and Condition.

2.2. Data quality concept and assessment methods

In this study, three DQ dimensions were considered, namely completeness, plausibility, and uniqueness - dimensions defined as most relevant together with domain experts in CORD-MI [6]. Five DQ parameters (parameters relevant for DQ but not indicating DQ) and seven DQ indicators are derived from these dimensions as shown in tables 1 and 2.

RD-specific coding with so-called Orphacodes (OCs) is necessary to avoid any ambiguity in RD documentation. Specific metrics are therefore used to assess the quality of RD data. For definitions of used DQ metrics please refer to [6].

The DQ library (dqLib) [16] has been used to develop specific reporting scripts for DQ assessments. This R package provides methods that enable users to select desired dimensions, indicators, and parameters as well as to define specific DQ reports [6]. Using this software framework, a specific DQ tool was implemented to import the employed data sets and configure required DQ reports [11]. To compare the quality of RD data before and after ETL processes, we applied the developed tools on the two data sets. The generated DQ report comprises two Excel spreadsheets for each data set. The first sheet illustrates the calculated DQ metrics as shown in tables 1 and 2, while the second sheet reports the detected DQ issues. All reports are checked manually for contradictions. If no contradiction can be found the reports are considered as correct.

3. Results

Table 1. DQ parameters displayed in the generated reports for 2020. The resulting RD cases are unambiguously identified by OCs or tracer ICD-10-GM codes. All rel. frequencies are normalized to 100.000 cases.

Data Set	Inpatient Cases	Analyzed Inpatient Cases	RD Cases rel. Frequency	Orpha Cases rel. Frequency	Tracer Cases rel. Frequency
SAP	79810	1415	649	538	241
FHIR	79810	1417	221	0	221

Table 2. DQ indicators displayed in the generated reports for 2020.

DQ Dimension	DQ Indicator	Abr.	SAP Data	FHIR Data
Completeness (co)	Item Completeness Rate	dqi_co_icr	85,71%	78,57%
	Value Completeness Rate	dqi_co_vcr	99,48%	95,64%
	Orphacoding Completeness Rate	dqi_co_ocr	53.73%	0
Plausibility (pl)	Orphacoding Plausibility Rate	dqi_pl_opr	93,88%	NA
	Range Plausibility Rate	dqi_pl_rpr	100%	100%
Uniqueness (un)	RD Case Unambiguity Rate	dqi_un_cur	94,02%	93,18%
	RD Case Dissimilarity Rate	dqi_un_cdr	50%	100%

Tables 1 and 2 present the results of DQ assessments performed on the employed data sets. Discrepancies between the DQ results obtained before and after ETL can be observed in most parameters and indicators. The FHIR data has no Orpha-coded cases and as a consequence less RD cases than the SAP data as illustrated in table 1. Table 2 shows that both indicators for item and value completeness are notably higher in SAP data than that in FHIR data. Furthermore, the unambiguity rate of RD cases is slightly higher in SAP data than in the FHIR data. However, the dissimilarity indicator achieved better results on FHIR data and reached its maximal level after the ETL processes.

4. Discussion

Table 1 shows a discrepancy between the parameters obtained before and after the ETL processes. Both data sets have the same origin (SAP IS-H) and show the same number of inpatient cases. However, the data in the PAS is not static but underlies corrections even on historical data. This may explain the little discrepancy found in Analyzed

Inpatient Cases and Tracer Cases rel. Frequency. It is a bit counterintuitive that the analyzed inpatient cases are higher in FHIR data, but the tracer cases are lower. However, as the German remuneration act allows for retrospective clarifications, recording in the PAS may still occur for previous years. Furthermore, the FHIR data does not contain any data items for capturing OCs, because the §21 Act did not support this standard in 2020 [14]. Only diagnoses with an unambiguous ICD-10-GM code (so-called tracer) allow the identification of RD cases. The legislature has responded to the necessity of better RD documentation and from April 2023 on, adding OCs according to Alpha-ID-SE terminology [17] will become mandatory for all RD documentation in §21 data.

Table 2 shows that the completeness indicators performed better on SAP data than on FHIR data. The main reason for that are FHIR mapping errors that have been identified using the DQ reports and will be therefore removed in the next update. Another reason is the lack of required OCs in the FHIR data as mentioned above. In contrast, the dissimilarity indicator performed better on FHIR data. This indicator shows that the SAP data contains duplicated RD cases. Using ETL the data were cleansed from duplications and transformed into a standardized format. Therefore, the dissimilarity indicator reached its maximal level after ETL processes. In addition, table 2 shows that although the FHIR data do not contain any OCs, our methods were able to compute the RD Case Unambiguity Rate only based on available ICD-10-GM codes. The ETL did not introduce any issues regarding the range plausibility.

The execution of DQ assessments runs without errors. There are no contradictions in the generated DQ reports. The study results have therefore indicated the correctness of calculated DQ metrics and detected DQ issues. The implemented metrics cover independent aspects of DQ [6]. Our study has shown that the developed methodology is capable of detecting potential DQ issues such as missings, implausibility or ambiguity of RD diagnoses and that it can be used for reporting on the quality of real-world data stored in heterogeneous data sources. The DQ reports helped us to compare the DQ before and after ETL processes and to find the causes of detected DQ violations. The results were validated independently by domain experts. The used DQ dimensions and metrics fit well the specified requirements, with certain limitations as described in [6]. In future works we will apply our methods on data stored in distributed data sources across multiple hospitals to compare the quality of RD data recorded in different HISs.

5. Conclusion

The resulting DQ reports have shown the correctness of calculated DQ metrics and detected DQ issues. Our work is the first study to investigate the impact of ETL processes on the quality of RD data to our knowledge. We showed that our methodology is able to identify DQ issues by comparing the DQ before and after ETL processes. We found discrepancies between the DQ results obtained before and after ETL processes, some of them based on errors in the transformation step. Such complex and challenging processes can decrease the quality of RD data. On the other hand, our study has shown that the extracted data were cleansed from duplications and transformed into an interoperable standard using ETL processes. This enhances the reuse of RD data for clinical research. We have demonstrated the usefulness and portability of developed tools by applying our methodology to real-world data stored in different data formats and structures. Our methodology can therefore be used to improve the quality of RD documentation and to support clinical research.

Ethical Approval: This study was performed in line with the principles of the Declaration of Helsinki. The analysis of retrospective, pseudonymized data collected for patients with rare disease diagnosis at University Hospital Tübingen in 2020, using the technical infrastructure of the MI-initiative, was approved by the ethics committee of University Hospital Tübingen (reference number 514/2020BO2).

Author Contributions: Conceptualization, methodology, and writing-original draft: KT and DK; requirement analysis and definition of indicators: KT, DK, TM, and HG; software for data quality assessments: KT; use case execution and software for data curation: KT and RV; writing-review and editing: KT, RV, DK, TM, and HG.

Acknowledgement: This study was done within the CORD-MI project funded by the German Federal Ministry of Education and Research (BMBF), funding numbers 01ZZ1911R and 01ZZ1911O.

References

[1] 'Use Case CORD-MI | Medizininformatik-Initiative'. https://www.medizininformatik-initiative.de/de/CORD (accessed Jun. 10, 2023).
[2] Semler SC, Wissing F, Heyder R. German medical informatics initiative. Methods of information in medicine. 2018 May;57(S 01):e50-6. doi: 10.3414/ME18-03-0003.
[3] Martin T et al., 'Problems of finding rare diseases in the documentation of German hospitals', Sep. 2021, p. DocAbstr. 59. doi: 10.3205/21gmds117.
[4] Tahar K, et al. Integrating heterogeneous data sources for cross-institutional data sharing: requirements elicitation and management in SMITH. InMEDINFO 2019: Health and Wellbeing e-Networks for All 2019 (pp. 1785-1786). IOS Press., doi: 10.3233/SHTI190647.
[5] Spengler H, et al. Improving data quality in medical research: a monitoring architecture for clinical and translational data warehouses. In2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) 2020 Jul 28 (pp. 415-420). IEEE. doi: 10.1109/CBMS49503.2020.00085.
[6] Tahar K. et al., 'Rare Diseases in Hospital Information Systems – An Interoperable Methodology for Distributed Data Quality Assessments', *Methods Inf Med*, vol. 0, no. AAM, doi: 10.1055/a-2006-1018.
[7] Ramasamy A. and Chowdhury S. 'Big Data Quality Dimensions: A Systematic Literature Review', J. Inf. Technol. Manag., vol. 17, no. 0, Art. no. 0, May 2020, doi: 10.4301/S1807-1775202017003.
[8] Schmidt CO, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Medical Research Methodology. 2021 Dec;21(1):1-5. doi: 10.1186/s12874-021-01252-7.
[9] 'SAP Software Solutions | Business Applications and Technology'. https://www.sap.com/index.html (accessed Dec. 30, 2022).
[10] 'BfArM - ICD-10-GM'. https://www.bfarm.de/EN/Code-systems/Classifications/ICD/ICD-10-GM/_node.html (accessed Jan. 10, 2023).
[11] Tahar K., Data Quality Assessment on Rare Diseases Data – A set of Metrics and Tools for the 33rd Medical Informatics Europe Conference 2023. https://doi.org/21.11101/0000-0007-FB71-F
[12] 'The MII core data set'. https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set (accessed Jan. 10, 2023).
[13] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. InProceedings of the 26th IEEE international symposium on computer-based medical systems 2013 Jun 20 (pp. 326-331). IEEE. doi: 10.1109/CBMS.2013.6627810.
[14] 'Datenübermittlung an das InEK.' https://www.dkgev.de/themen/digitalisierung-daten/elektronische-datenuebermittlung/datenuebermittlung-an-das-inek/ (accessed Dec. 29, 2022).
[15] 'CORD-MI ImplementationGuide'. https://simplifier.net/guide/medicalinformaticsinitiative-cord-implementationguide (accessed Dec. 23, 2022).
[16] Tahar T, Data Quality Library (dqLib): R package for data quality assessment and reporting. 2022. https://doi.org/21.11101/0000-0007-F6DE-A (accessed Dec. 29, 2022).
[17] 'BfArM Alpha-ID-SE'. https://www.bfarm.de/EN/Code-systems/Terminologies/Alpha-ID-SE/_node.html (accessed Jan. 10, 2023).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230122

Agent Based Modelling for Simulating the Interregional Patient Mobility in Italy

Fabrizio PECORARO[a,1], Filippo ACCORDINO[a], Federico CECCONI[b] and Mario PAOLUCCI[a,b]
[a] *Institute for Research on Population and Social Policies, National Research Council. Rome, Italy*
[b] *Institute of Cognitive Sciences and Technologies, National Research Council. Rome, Italy*
ORCiD ID: Fabrizio Pecoraro https://orcid.org/0000-0001-5718-4240, Filippo Accordino https://orcid.org/0000-0002-4245-0654, Federico Cecconi https://orcid.org/0000-0003-3336-775X, Mario Paolucci https://orcid.org/0000-0002-8276-1086

Abstract. Patient mobility is considered one of the main concerns for policy-makers as it impacts financial sustainability of regional health systems due to the high percentage of patients accessing care services in other regions. For a better understanding of this phenomenon, it is necessary to define a behavioral model able to represent the patient-system interaction. In this paper we adopted the Agent-Based Modelling (ABM) approach with the aim of simulating patient flow across regions and determining which are the main factors influencing it. This may provide a new insight for policy makers to capture which are the main factors influencing mobility and actions that may contribute to contain this phenomenon.

Keywords. Patient mobility, Agent-Based Modelling, Italy, Spatial accessibility, Simulation process

1. Introduction

Patient mobility studies the migration of patients to access to health services located outside his/her region of residence. It is considered as a proxy to appraise the quality and availability of hospital services [1] and to point out socio-economic disparities at local and regional level [2]. Different studies have underlined the main factors influencing this phenomenon [1], such as demographic and socio-economic status [3], quality and complexity of local services [4], structural components [5]. Among structural factors two inter-related aspects affect patient mobility: accessibility and availability of intra- and inter-regional facilities in particular for patients living at the regional borders [6]. In Italy, mitigating passive mobility is one of the main actions at the basis of the Health Pact 2019–2021 signed by the Conference of Regions and Autonomous Provinces [7]. This Health Pact highlights the necessity of mapping patient flows and drawing up a "plan to stop" passive mobility, with particular attention on critical care [1]. To further understand this phenomenon and capture which are the

[1] Fabrizio Pecoraro, IRPPS-CNR, Via Palestro, 32, 00185 Rome, Italy E-mail: f.pecoraro@irpps.cnr.it

main factors influencing patient's choice, it is necessary to define a behavioral model that includes a variety of individual, community and socio-economic characteristics that represent the patient-system interaction. In different complex settings among which healthcare, this is done using Agent-Based Modelling (ABM) approach [8]. ABM is able not only to synthesize prior knowledge of a population and effectively represent and simulate this interaction [9], but also to understand how an intervention could modify the dynamics of patient behavior and affect public health [10]. Within this context, aim of this study is to design and propose an ABM approach for simulating patient flow across the Italian regions and determining which are the main factors influencing it. The suitability of this approach in this specific setting is tested considering both the accuracy and precision of the simulation process. The application of a robust ABM allows to define a mathematical model describing patient flow at an abstract level (i.e., region) and apply it to simulate the patient's behavior at a refined level (i.e., municipality). Moreover, this approach may provide a new insight for policy makers to capture which are the main factors influencing inter-regional patient mobility and to identify and put in place actions that may contribute to contain this phenomenon.

2. Materials and Methods

2.1. Data collection and identification of factors

Data on hospitals and mobility were gathered from the Ministry of Health website [11] and from the National Outcomes Program website [12], while demographic data were collected from the Italian National Institute of Statistics (ISTAT) website [13]. All data refers to the year 2019. From a clinical perspective, this study focuses on the hip replacement surgery procedure, an elective treatment where patients are generally prepared to travel long distances beyond their nearest provider [14]. To identify which variables mostly impact patient mobility, we applied the best subsets regression function of R (i.e., *regsubsets*) that tests all possible combinations of the predictor variables and then selects the best model according to the highest adjusted R squared value. The resultant regression model is reported in Equation 1 (note that adjusted R^2 = 0.66 and all variables are statistically significant, $p < 0.05$).

$$stay_{\%} = 42 - 0.05w + 0.6s + 0.07int_{intra} - 50ret_{intra} + 0.3bed_{intra} \quad (1)$$

where $stay_{\%}$ is the probability that the patient is cared in his/her region of residence, w is the number of waiting days to access the service (at regional level), s is the level of patient satisfaction due to the last hospital admission (at regional level), while int_{intra}, ret_{intra} and bed_{intra} describe, respectively, the number of interventions, the percentage of patients returned to hospital in the following 2 years from the intervention and the number of beds available in the orthopedics wards. Note that further indicators such as those related to the patient (i.e., income, education) have been discarded from the model as they were not statistically significant. Hospital-related indicators has been computed (for each municipality i) using Equation 2 based on a gravity model which relates the probability to access to a hospital with its capacity, quality and distance:

$$int^{i}_{intra} = \sum_{j\in\{Reg(i)=Reg(j)\}} INT_j R_j w_{i,j} = \sum_{j\in\{Reg(i)=Reg(j)\}} \frac{INT_j}{\sum_i pop_i * w_{i,j}} w_{i,j} \quad (2)$$

where R_j represents the weighted hospital-to-population index of hospital j, INT_j is the number of interventions carried out in hospital j and pop_i is the resident population of the municipality i. $w_{i,j}$ that represents the weighting distance between the hospital j

and the municipality *i* has been computed using the Sigmoid decay function. Based on Equation 2, an average value weighted by population of each indicator has been computed at province level. For further details please see [15].

2.2. Simulation process description

Figure 1 shows the main steps of the ABM simulation process. A set of 100.000 individuals are extracted from the whole target population using a stratified random sampling methodology considering two risk factors: age and gender. The second step (i.e., simulation) is composed by three activities: 1) a set of 1000 patients are extracted from the sample population to define the group of patients to be cared in a specific loop; 2) selected patients and hospitals, both represented by turtles, are placed over the patches of the environment based on their belonging municipality; 3) one patient at time is randomly picked up from the 1000 patients and the staying index (see Equation 1) is computed to assess the probability that the patient remains in his/her region to be cared. This index has been also applied to capture the level of attractiveness of each hospital located both within and outside the patient region of residence. Finally, the ABM simulation process stochastically determines the healthcare structure chosen by the patient. This process has been executed 52 times to simulate the access to care as a weekly procedure considering that the average length of stay for the primary total hip replacement is around 7 days [16].

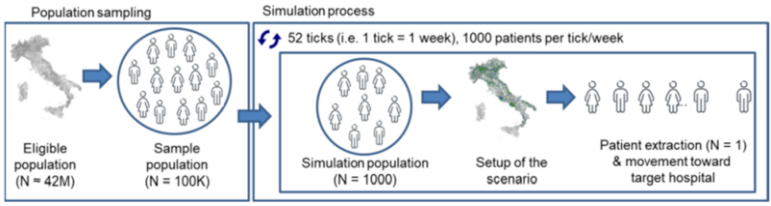


Figure 1. Agent-based modelling simulation process

To capture the accuracy (i.e., reproducibility) and the precision (i.e. repeatability) of the model five repeated sessions has been executed. From a statistical perspective, accuracy was assessed using the regression coefficient (R) between the simulated data and the original data, while precision was assessed using the Intraclass Correlation Coefficient (ICC(2,1)) to capture the intra-session reliability between observations.

3. Results

Figure 2 shows the Netlogo environment: on the left side the interface items adopted to control agents and the system are reported. In particular, the *n_ticks* and *n_patients* inputs allow to set, respectively, the number of weeks and the number of patients per week to be involved in the simulation. The output *patients_to_be_placed* facilitates the supervision of the process status capturing the number of patients located on the environment that has not been cared yet. On the right side of the window Netlogo integrates the map of the Italian territory divided by municipalities, colored depending on the passive mobility percentage, as reported in the specific legend.

The scatterplot diagram shown in Figure 3 highlights the linear regression between the passive mobility gathered from the ABM simulation (x-axis) and the passive mobility computed with the multiple linear regression model (y-axis). As clearly

reported by R^2 (= 0.94) there is a very strong direct relationship between these two variables. A high correlation (R^2 = 0.66) is also reported considering the linear regression between the simulation mobility and passive mobility computed with the real hospital values. This result confirms the goodness of the simulation model. Considering the precision, the ICC(2,1) computed carrying out five sessions of ABM simulation resulted higher than 0.95 confirming the repeatability of the process.

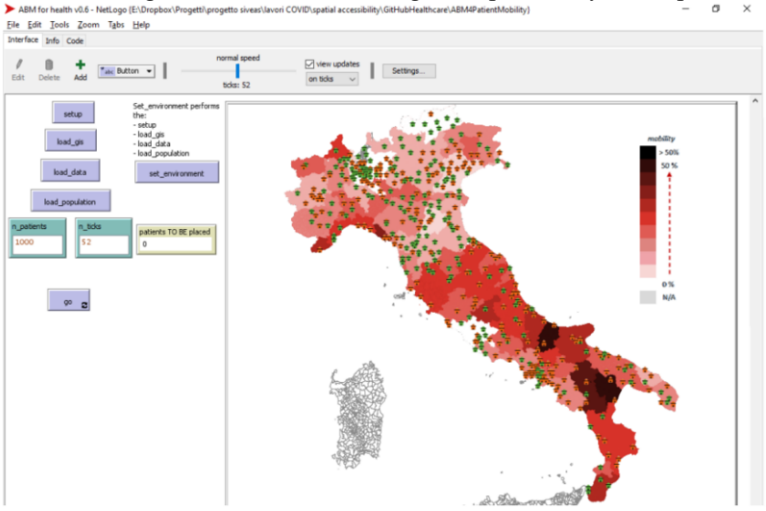


Figure 2. Netlogo environment highlighting the preliminary results of one simulation session

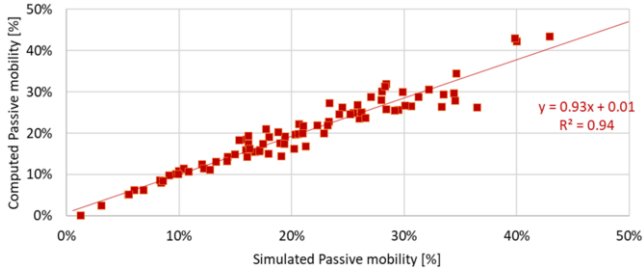


Figure 3. Correlation between the passive mobility gathered from the simulation model (x-axis) and the passive mobility computed with the multiple linear regression model (y-axix)

4. Discussion and Conclusions

The aim of this paper is twofold: on the one hand it would like to provide an ABM approach able to accurately simulate patient flow across the Italian regions and, on the other hand to determine which are the main individual, hospital and local factors influencing patient's mobility. To accomplish this task, we firstly defined a mathematical model able to accurately describe the dynamics of the patient-system interaction and define the probability that each patient involved in the simulation process accesses to an inter-regional structure. Based on this model, ABM stochastically determines the level of attraction of each structure and simulates the patient flows across the Italian regions. Preliminary results of the applicability of this approach highlight a high precision and accuracy in the description of patient mobility. This is clearly evident considering the very strong correlation coefficient between the

simulated and the computed passive mobility. Moreover, the repeatability of the process is also confirmed by analyzing the low inter-session variability reported by the ICC. This high level of accuracy and precision confirms the goodness of the ABM simulation approach to describe this specific scenario. In this paper, we applied this methodology considering the hip replacement surgery procedure. However, this can also be applied to other elective surgery or curative services, to primary care services, or even to acute care services, such as intensive care.

The results reported in this paper focus on a real patient passive mobility scenario based on a prior knowledge of a population. Future works will detail the mathematical model to further analyze the main factors at individual, local and regional level responsible for attracting patients and contribute to active mobility. Moreover, simulation basic variables will be updated to verify how these changes may impact on patient mobility. This may help policy makers and hospital administrative professionals to capture to what extent these changes may help to contain these dynamics. For instance, this can be done by reducing the waiting times or improving the availability of services in a specific part of the territory that is not reached by the service under investigation. This can be done either by improving beds, personnel and resources on already established facilities or even providing an additional point of care.

References

[1] Nante N, Guarducci G, Lorenzini C, Messina G, Carle F, Carbone S, Urbani A. Inter-Regional Hospital Patients' Mobility in Italy. Healthcare. 2021; 9(9): 1182.

[2] Ricci A, Barzan E, Longo F. How to identify the drivers of patient inter-regional mobility in beveridgean systems? Critical review and assessment matrix for policy design & managerial interventions. Health Services Management Research. 2021; 34(4): 258-268.

[3] Rubino C, Di Maria C, Abbruzzo A, Ferrante M. Socio-economic inequality, interregional mobility and mortality among cancer patients: A mediation analysis approach. Socio-Economic Planning Sciences. 2022; 82: 101247.

[4] De Curtis M, Bortolan F, Diliberto D, Villani L. Pediatric interregional healthcare mobility in Italy. Italian Journal of Pediatrics. 2021; 47(1): 1-7.

[5] Balia S, Brau R, Marrocu E. Interregional patient mobility in a decentralized healthcare system. Regional Studies. 2018; 52(3): 388-402.

[6] Pecoraro F, Luzi D, Clemente F. Spatial Inequity in Access to Intensive Care Unit Beds at Regional Level in Italy. Studies in Health Technology and Informatics. 2021; 281: 809-813.

[7] Italian State-Regions Standing Conference. Health Pact 2014-2016. 2014. Available in Italian at: https://www.salute.gov.it/imgs/C_17_pagineAree_2986_listaFile_itemName_8_file.pdf

[8] Tracy M, Cerdá M, Keyes KM. Agent-Based Modeling in Public Health: Current Applications and Future Directions. Annu Rev Public Health. 2018; 39: 77–94.

[9] Steinbacher M, Raddant M, Karimi F. et al. Advances in the agent-based modeling of economic and social behavior. SN Bus Econ. 2021; 99.

[10] Marshall B. Agent-based modelling. In: El-Sayed AM, Galea S, editors. Systems science and population health. Oxford: Oxford University Press; 2017. p. 29–51.

[11] Ministry of Health. National healthcare service database. 2022. Available from: www.salute.gov.it/portale/documentazione/p6_2_8_1_1.jsp?lingua=italiano&id=6

[12] National Agencies for Regional Health Services (AGENAS). National Healthcare Outcomes Program. Available. 2022. from: https://pne.agenas.it/

[13] Italian National Institute of Statistics (ISTAT). Available from: https://www.istat.it/it/archivio/222527

[14] Berta P, Guerriero C, Levaggi R. Hospitals' strategic behaviours and patient mobility: Evidence from Italy. Socio-Economic Planning Sciences. 2021: 101030.

[15] Pecoraro, F., Luzi, D., & Clemente, F. (2022). The Impact of Hospital Accessibility on Interregional Patient Mobility in Italy. Studies in Health Technology and Informatics, 294, 684-688.

[16] Foote, J., Panchoo, K., Blair, P., & Bannister, G. (2009). Length of stay following primary total hip replacement. The Annals of The Royal College of Surgeons of England, 91(6), 500-504.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230123

Towards a Consistent Representation of Contradictions Within Health Data for Efficient Implementation of Data Quality Assessments

Khalid O. YUSUF[a], Sabine HANSS[a,c] and Dagmar KREFTING[a,b,c,1]
[a] *Department of Medical Informatics, University Medical Center Göttingen, Göttingen, Germany*
[b] *Campus Institute Data Science (CIDAS), Georg-August-University, Göttingen, Germany*
[c] *German Center for Cardiovascular Research, Partner Site Göttingen, Germany*
ORCiD ID: Khalid O. Yusuf https://orcid.org/0000-0003-2556-6898

Abstract. Contradictions as a data quality indicator are typically understood as impossible combinations of values in interdependent data items. While the handling of a single dependency between two data items is well established, for more complex interdependencies, there is not yet a common notation or structured evaluation method established to our knowledge. For the definition of such contradictions, specific biomedical domain knowledge is required, while informatics domain knowledge is responsible for the efficient implementation in assessment tools. We propose a notation of contradiction patterns that reflects the provided and required information by the different domains. We consider three parameters (α, β, θ): the number of interdependent items as α, the number of contradictory dependencies defined by domain experts as β, and the minimal number of required Boolean rules to assess these contradictions as θ. Inspection of the contradiction patterns in existing R packages for data quality assessments shows that all six examined packages implement the (2,1,1) class. We investigate more complex contradiction patterns in the biobank and COVID-19 domains showing that the minimum number of Boolean rules might be significantly lower than the number of described contradictions. While there might be a different number of contradictions formulated by the domain experts, we are confident that such a notation and structured analysis of the contradiction patterns helps to handle the complexity of multidimensional interdependencies within health data sets. A structured classification of contradiction checks will allow scoping of different contradiction patterns across multiple domains and effectively support the implementation of a generalized contradiction assessment framework.

Keywords. Data quality, health data, boolean minimization, metadata

[1] Department of Medical Informatics, University Medical Center Göttingen, Germany; E-mail: dagmar.krefting@med.uni-goettingen.de.

1. Introduction

Contradictions in data quality (DQ) assessments are typically understood as (nearly) impossible combinations of data values of interdependent data items within a data set [1]. In health data, a typical example is that diastolic blood pressure (DP) must not be higher than systolic blood pressure (SP). Handling of a single interdependency between two data items is well established and implemented in most DQ assessment frameworks, as we show below. There are several taxonomies used in literature to describe the semantic nature (e.g. logical or empirical) of contradictions [2–5]. This is useful for the definition of contradictions on the biomedical domain expert level, while the implementation within a DQ assessment tool is typically realized by Boolean rules that are somehow agnostic to the semantic nature of a contradiction. For example, a conditional expression ($x > y$) might be related (a) to DP and BP as well as (b) age at two different time points. While these examples might be assigned to different semantic groups, for example (a) as atemporal and (b) as temporal plausibility according to [4], from the implementation point of view, a clear and simple implementation of contradiction rules is desired [6]. Therefore, we propose a structural representation of contradiction patterns that is useful to simplify increasing complexity of multidimensional interdependencies, as it describes the dimensionality of the contradictory dependency and the minimum of derived Boolean rules. For the aforementioned blood pressure case, there is only one contradictory dependency between the two data items. Another pair of data items however may have more interdependencies. An example is the notation of fever and the body temperature: We have to distinguish the two cases where the body temperature is below or above a certain temperature, where one case requires the fever_item set to *yes*, while the other requires the fever_item set to *no* or at least *not set*. This would result in two Boolean rules in the implementation.

2. Methods

We consider three parameters (α, β, θ) for the proposed structural representation of contradictions i.e. the number of: 1) interdependent data items α, 2) contradictory dependencies β, and 3) minimal Boolean rules θ. While β represents the number of distinct contradictions defined by biomedical domain experts, θ is derived by grouping multiple similar contradictions using all plausible common denominators (CD). A CD could either be a conditional expression, an item, a value or their combination. As a rule, any defined rule that evaluates to multiple numerical or categorical values is set to its value-range or value-set respectively—provided it represents a distinct contradiction. Also, each minimal Boolean rule within θ must be bounded unambiguously such that it is independent of other rules. We examined six R-packages that support contradiction assessment *(assertive, dataquierR, DQAstats, pointblank, testdat, and validate)* on what Boolean rules are implemented in these packages [7]. Based on recent quality assessments on biobank and COVID-19 data [8,9], we defined different classes of contradiction patterns on the sets of interdependent data items.

3. Results

3.1. Contradiction pattern implemented in R-packages

As shown in table 1, all R-packages implemented contradiction checks on $\alpha = 2$ with $\beta = 1$ which translates to $\theta = 1$ in each case. These scenarios represent the simplest form of contradiction patterns.

Table 1. Contradiction pattern in existing R-packages (c.f. [7] for information on R-packages). α = number of interdependent items, β = number of contradictory dependencies, θ = number of Boolean rules

R-Package	Notation (α, β, θ)	α	β	θ
assertive	*(2,1,1)*	x,y	Is_less_than(x,y)	1
dqastats	*(2,1,1)*	bank_balance(bb), credit_worthiness(cw)	is_negative(bb) & cw == yes	1
pointblank	*(2,1,1)*	x,y	col_val_lt (vars(x), y)	1
testdat	*(2,1,1)*	x,y	expect_cond(x, y.length>=1)	1
validate	*(2,1,1)*	staff, staff_cost	validator(if (staff >=1) staff_cost >= 1)	1
con_contradictions	*(2,1,1)*	Age_followup, Age_baseline	A_less_than_B(Age_1, Age_0)	1

3.2. Contradiction pattern in the biobank domain

Three interdependent data items used in storing information about the collection of citrate samples were investigated, i.e. primary receptables (pr), desired aliquots filled (af), and actual aliquots count (ac) [8]. From table 2, a contradiction pattern was established where $\alpha = 3$, $\beta = 5$, and $\theta = 3$. Transforming β to θ resulted in the reduction of 4 distinct contradictory rules to 2 Boolean rules using their respective CD (pr==0) and (pr > 0 & af == no & ac ==). Another contradiction pattern refers to contradictions in items related to pre-analytic states of blood samples, please refer to [8] for further information.

Table 2. Contradiction pattern between biosample associated data items

α	β	θ
Number of primary receptable (pr)	pr == 0 & af == yes	pr == 0 & isTrue(af == yes | ac > 0)
All aliquots filled (af)	pr == 0 & ac > 0	pr > 0 & af == no & ac in (0,4)
Aliquot count (ac)	pr > 0 & af == no & ac == 0 pr > 0 & af == yes & ac < 4 pr > 0 & af == no & ac == 4	pr > 0 & af == yes & ac < 4

3.3. Contradiction pattern in the COVID-19 domain

We considered the consistency of different groups of interdependent data items that were mapped from different cohorts to the German Corona Consensus (GECCO) dataset [9]. Table 3 shows the aforementioned contradiction pattern between fever and body temperature (T_b) items: a normal T_b ($35 \leq T_b < 38.3$) should not evaluate to fever_item set to *yes* and an elevated T_b ($38.3 \leq T_b \leq 45$) should not evaluate to fever_item set to *no* or *not set*. There are no plausible CD in this case, hence, β equals θ and is assigned to

class (2,2,2). The T_b ranges and fever_item being compared to the elevated T_b evaluate to multiple values pointing to distinct contradictions.

Table 3. Contradiction pattern between body_temperature (T_b) and fever

α	β	θ
fever body_temperature (T_b)	35 ≤T_b<38.3 & fever == yes 38.3≤T_b≤45 & fever in (no, notset)	T_b>=35 & T_b<38.3 & fever == yes T_b>=38.3 & T_b<=45 & fever in (no, notset)

A more complicated example is the *(10,10,2)* class for a comparison on the presence of pulmonary disease (PD) anamnesis with its nine documented indicators. PD is a dependent variable on its set of indicators such that when a study participant answers the question about the presence of PD affirmative, this has to be followed-up with at least a specific disease that belongs to the PD family. As presented in table 4, β is derived from the combination of values of PD and their implausible indicators. By exploiting the CD, β is reduced to θ = 2. While β preserve the atomicity of the rules provided by domain experts, θ is implemented in the assessment tool by grouping multiple similar β within θ. Further examples are the comparison of COVID-19 severity and 19 severity indicators, resulting in a (20,38,4) pattern (c.f. [9] for further information).

Table 4. Contradiction pattern between branch question of pulmonary disease anamnesis and documented indicators.

α	β	θ
Chronic Lung Disease (pd)?	pd == no & asthma == yes	pd == no & any_of(asthma, copd, fibr, ph, ohs, apn, osas, cf, others) == yes
Asthma	pd == no & copd == yes	pd == yes & all_of(asthma, copd, fibr, ph, ohs, apn, osas, cf, others) == no
Chronic obstructive pulmonary disease (copd)	pd == no & fibr == yes	
Lung fibrosis (fibr)	pd == no & ph == yes	
Pulmonary hypertension (ph)	pd == no & ohs == yes	
Obesity hypoventilation syndrome (ohs)	pd == no & apn == yes	
Sleep apnoea (apn)	pd == no & osas == yes	
Obstructuve sleep apnoea (osas)	pd == no & cf == yes	
Cystic fibrosis (cf)	pd == no & others == yes	
Other lung disease (others)	pd == yes & all_pd_indicators == no	

4. Discussion

We demonstrate an efficient way of representing the dimensionality of contradictions where the underlying structure of multiple rule combinations is described. Our results indicate θ may be significantly lower than β in multidimensional interdependencies. While the preservation of β will aid the explanation and traceability of identified contradictions, θ ensures an efficient implementation of Boolean rules within DQ

assessment tools—in particular for large datasets. Though different contradiction indicators are useful for the semantic description of contradictions, a structural classification simplifies the varying dimensionality of contradictions for ease of evaluation. A holistic approach to contradiction assessment is a step towards building a DQ assessment tool that would be applicable across multiple domains. Uniform representation of contradiction patterns will ensure a harmonized way of comparing contradiction patterns when considering the fitness of existing assessment tools for internal use. Conclusively, a structured classification of contradiction checks will support the implementation of a generalized contradiction assessment framework effectively and may help researchers to identify Boolean rules in a structured way. While the evaluations in the described cases are performed manually, we envision a tool that helps to translate the domain specific definition of contradictory rules to a normalized form of the informatics domain.

Acknowledgement: The work is jointly funded by NUM, grant number 01KX2121 and DZHK, grant number 81Z0300108

Conflict of Interest: The authors declare, that there is no conflict of interest.

Ethical approval and consent: DZHK and TORCH approved the use of the TORCH dataset. Also, use of the data from the cohorts was approved by the NAPKON use and access committee.

References

[1] De Marneffe MC, Rafferty AN, Manning CD. Finding contradictions in text. InProceedings of acl-08: Hlt 2008 Jun (pp. 1039-1047).

[2] Schmidt CO, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Med Res Methodol. 2021 Dec;21(1):63. doi: 10.1186/s12874-021-01252-7

[3] Johnson SG, Pruinelli L, Hoff A, Kumar V, Simon GJ, Steinbach M, et al. A Framework for Visualizing Data Quality for Predictive Models and Clinical Quality Measures. AMIA Jt Summits Transl Sci Proc. 2019;2019:630–8

[4] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMs. 2016 Sep 11;4(1):18. doi:10.13063/2327-9214.1244

[5] Nonnemacher M, Nasseh D, Stausberg J, Bauer U. Datenqualität in der medizinischen Forschung: Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. 2., aktualisierte und erw. Aufl. Berlin: Med. Wiss. Verl.- Ges; 2014. 230 p. (Schriftenreihe der TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V).

[6] Duşa A. A mathematical approach to the boolean minimization problem. Qual Quant. 2010 Jan;44(1):99–113. doi:10.1007/s11135-008-9183-x

[7] Mariño J, Kasbohm E, Struckmann S, Kapsner LA, Schmidt CO. R Packages for Data Quality Assessments and Data Monitoring: A Software Scoping Review with Recommendations for Future Developments. Applied Sciences. 2022 Apr 22;12(9):4238. doi:10.3390/app12094238

[8] Yusuf K, Tahar K, Sax U, Hoffmann W, Krefting D. Assessment of the Consistency of Categorical Features Within the DZHK Biobanking Basic Set. In: Röhrig R, Grabe N, Hoffmann VS, Hübner U, König J, Sax U, et al., editors. Studies in Health Technology and Informatics [Internet]. IOS Press; 2022 [cited 2022 Sep 10]. Available from: https://ebooks.iospress.nl/doi/10.3233/SHTI220809

[9] Yusuf KO, Miljukov O, Hanß S, Schoneberg A, Wiesenfeldt M, Stecher M, et al. Consistency as a Data Quality Measure for German Corona Consensus items mapped from National Pandemic Cohort Network data collections. Methods Inf Med. 2023 Jan 3. doi:10.1055/a-2006-1086

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230124

Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned

Hans-Ulrich Prokosch[a,1], Marie Gebhardt[b], Julian Gruendner[a], Philip Kleinert[b], Karoline Buckow[b], Lorenz Rosenau[c] and Sebastian C. Semler[b]
[a]*Chair of Medical Informatics, University of Erlangen-Nürnberg, Erlangen, Germany*
[b]*TMF - Technology, Methods, and Infrastructure for Networked Medical Research, Berlin, Germany*
[c] *IT Center for Clinical Research, University of Lübeck, Lübeck, Germany*

Abstract. Harmonizing medical data sharing frameworks is challenging. Data collection and formats follow local solutions in individual hospitals; thus, interoperability is not guaranteed. The German Medical Informatics Initiative (MII) aims to provide a Germany-wide, federated, large-scale data sharing network. In the last five years, numerous efforts have been successfully completed to implement the regulatory framework and software components for securely interacting with decentralized and centralized data sharing processes. 31 German university hospitals have today established local data integration centers that are connected to the central German Portal for Medical Research Data (FDPG). Here, we present milestones and associated major achievements of various MII working groups and subprojects which led to the current status. Further, we describe major obstacles and the lessons learned during its routine application in the last six months.

Keywords. Large scale data sharing, real world data, federated data network

1. Introduction

To leverage the power of real world data from all German university hospitals, the medical informatics initiative (MII) was initiated by the German Ministry of Research and Education [1]. Four consortia (DIFUTURE [2], HiGHmed [3], MIRACUM [4], and SMITH [5]) and a national coordination office [6] received funding from 2018 to 2022 to unlock the heterogeneous data silos across the German university hospital landscape and to enable large scale data sharing throughout Germany. Each consortium had independently defined its concept and practical approach towards data sharing within the partner sites of the consortium. Additionally, the MII national steering committee (NSC: comprising coordinators of each consortium, representatives from the national MII coordination office, and the Federal Ministry of Education and Research) was created as an overarching governance structure. Through these structures, a Germany-wide data sharing network across German university hospitals´ data integration centers (DIC) was

[1] Corresponding Author: Hans-Ulrich Prokosch, Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen, Germany; E-mail: Hans-Ulrich.Prokosch@fau.de.

established. The NSC initiated cross-consortia working groups (WGs) focused on data sharing concepts, patient consent, interoperability, and communication. The final aim of the MII was to support clinical and translational medical researchers by providing access to clinical patient data from all German university hospitals through one central entry portal. The portal enables researchers to:

- get a general overview of data and data types available across all German university hospitals within their local DIC,
- characterize their cohort of interest for planned research analysis,
- perform feasibility queries to retrieve the size of matching patient datasets,
- define data use proposals to request data from all integrated DIC and finally receive the proposed type of access for their research projects (e.g., central or federated analysis).

Such a portal (the German Portal for Medical Research Data = Deutsches Forschungsdatenportal für Gesundheit = FDPG) has been implemented and provided to the German university medicine research community as a beta-release in October 2022 and will be opened for general use in the first quarter of 2023.

The objective of this publication is to illustrate the general MII data sharing framework, the technical status of the FDPG, and the lessons learned from a projectathon in which the FDPG was used as data sharing framework.

2. Methods

Important milestones towards the FDPG have been achieved within the MII working groups “data sharing,” “consent,” and “interoperability,” which tackled the numerous harmonization/ standardization tasks required to align all German university hospitals in a joined, large scale data sharing network. Major results are published on the MII website [7]. The FDPG framework consists of three modules. First, a Germany-wide feasibility tool, which was developed and successfully deployed as a sub-project of (1) the CODEX project (COVID-19 data exchange platform: 2020/2021) [8] and (2) the MII- project ABIDE_MI (Aligning Biobanks and Data Integration Centers efficiently: 2021/2023) [9]. Second, The FDPG research proposal management module was developed by a commercial software development partner (Appsfactory GmbH). A third module, a transparency register was set up as a webpage designed to make research projects visible and understandable to patients. Parts of the FDPG middleware are based on consortia concepts and have been integrated into the FDPG ecosystem. The project partners have delivered central components for the portal and matching components for the decentral sites.

The decentral regulatory foundation and the local technical infrastructures were iteratively established at 31 German university hospitals within their DIC. This includes the establishment of associated data (and biosample) use and access committees (UACs) as well as local trusted third parties. To verify the practicability and performance of the FDPG, a MII-wide projectathon was pursued to test data sharing concepts and tools. Four research projects were prepared, submitted, and managed using the FDPG platform.

3. Results

3.1. NSC working group results

The regulatory framework for MII-wide data sharing has been described in UML-based process maps and associated legal documents, such as data/biosample use regulations, contracts, and the data use project proposal form. Results have been discussed and agreed upon in the cross-consortia WG "data sharing," shared with and consented to by the legal departments of all German university hospitals, and finally approved by the NSC.

To allow the use of patient data, documented primarily within hospitals´ patient care processes for medical research, a comprehensive discussion and consensus process has been initiated by the WG "consent" to define a template for a modular Broad Consent patient information/ consent form. After numerous rounds of discussion and revision, it was approved by all German state data protection officers and a WG of all German ethics committees in 2020 [10]. In summer 2021, German university hospitals have started "rollout projects" to implement the patient information and broad consent collection process into local workflows. About 100,000 patients have already signed this consent.

The MII-wide "information model" for a harmonized representation of data to be shared across all German DIC is one of the major achievements of the WG "interoperability." HL7 FHIR resources form the basis of the MII CDS. Until today, official versions of the six basic CDS modules - patient, encounter, diagnosis, procedure, laboratory data, and medication data, as well as three extension modules - consent, biosample, and intensive care medicine, have been developed, balloted, and approved by the German HL7 community. Implementation guides (IG) are available as simplifier IGs [11]. Data from 7.6 million patients are currently available, including data items for more than 152 million lab results, 85 million diagnoses, and 37 million procedures.

3.2. CODEX and ABIDE_MI results

The COVID-19 data exchange platform (CODEX) was initiated as a joint action of the four MII consortia to tackle the COVID-19 pandemic [8]. One of the platform components was a central feasibility portal accessing the federated FHIR servers in German university hospitals. The architecture and technology for those developments were based on the MII concepts and design principles supporting smooth scalability even within the more complex and comprehensive structures of the MII. Thus, the MII ABIDE_MI project extended the small underlying datasets from COVID-19 patients to encompass comprehensive clinical data from all hospital patients based on the six basic modules of the CDS, the consent module, and the biosample module.

In October 2022, 27 DIC were able to integrate their FHIR servers with routinely collected clinical data into the FDPG framework. Eight biobanks integrated biosample information into DIC FHIR servers. The comprehensive architecture of the feasibility tool, its middleware as well as decentralized local components has been described by Gruendner et al. [12] and Rosenau et al. [13]. The complete system was demonstrated at the MII symposium.

3.3. Projectathon results

New processes and tools for the infrastructure are regularly evaluated in MII-projectathons. The seventh MII-projectathon was dedicated to evaluating the regulatory

framework and the FDPG-tools in combination with tasks carried out by administrative staff, as well as the data preparation and delivery process during the execution of four real research projects (comprising federated as well as centralized analysis). 31 German university hospitals were requested to deliver data. 30 DIC responded to the proposed projects, with 21 DIC participating in at least one project. Overall 42,7 % (53 out of 124) responses among the four projects were positive. Thus, sufficient data could be provided for all projects. The main reasons for not providing data were the need for patient consent forms available and the missing implementation of MII CDS modules. All contracts were administered and distributed to participating sites. Currently, signatures are being collected. Data sharing will be initiated once projects are published in the transparency registry.

Regarding the application of the research proposal management module: we learned that too much manual rework and communication to applicants was still necessary by FDPG staff to finalize project proposals with the researchers. This is mainly due to the current lack of FAQs and information in the FDPG web interface. For instance, free text data entry fields for project proposals were unclear and required further clarification. Similar problems emerged regarding user interfaces designed for DIC and UACs. Implemented processes were still novel and required additional clarification by FDPG staff. Integrating decision-making processes into local DIC processes does function but will need to be further optimized in the future. A time critical issue is the laborious process of signing contracts, as not all university hospitals accept digital signatures. Paper-based signature collection – considering the large number of contract partners – is time consuming and represents a considerable burden for all involved. Further, free text descriptions of the required data were not specific enough, requiring clarification cycles between DIC, FDPG staff, and data requestor. In a future release of the FDPG framework, we will provide users with a hierarchical data catalogue, allowing the precise selection of the required data elements to avoid misunderstandings between the requestor and the data provider. We expect additional optimization requirements concerning the processing pipeline with respect to data preparation and delivery when all projects are finished. However, beyond all technical support, there will need to be further organizational process optimization and governance-level decisions to appropriately address the important non-technical issues in data sharing mediated through the FDPG.

4. Discussion and Outlook

The MII aims to share data across all German university hospitals for research. The MII Symposium and the seventh projectathon have shown that MII is close to this goal but far from being lean, fast, and efficient in its implementation of all processes, which is necessary to save operational resources and for enabling high-quality research. To date, due to the results of the CODEX and ABIDE_MI projects, the FDPG is able to give researchers an impression of what data they can request from the university hospitals. After conducting a feasibility analysis, researchers can write a project proposal, and the FDPG supports spreading this proposal to all participating DIC. Local processes have been established at all German university hospitals to review those proposals within their UACs and then accept or reject the proposed project and provide data. As a final regulatory step, a joint data use contract is signed by all data providing sites and the respective project coordinator based on the MII data use contract template. Data provided by the university hospitals for central data pooling and central data analysis were – in the

recent projectathon - collected by one DIC. The individual datasets were integrated into a single dataset and securely provided to the project coordinator. Those results illustrate the success of the presented approach. Most sites responded to the data requests with a high proportion of UACs approving data access. This showcases the acceptance of the processes, especially considering the novelty of this framework.

However, data harmonization across all universities remains challenging. The MII CDS serves as a good baseline but is insufficiently precise for the use-case of federated queries that requires strong data harmonization to fulfil performance requirements. Further restrictions on top of the MII CDS are therefore necessary to ensure interoperability, and shared tooling is required to guarantee data quality across sites. Integration of a FHIR terminology server proved essential to provide a hierarchic search ontology [13] and will in the future be crucial to verify data accuracy and quality. Further enhancements of the FDPG and extensions to the comprehensive FDPG ecosystem (to be implemented in central MII infrastructure projects of the MII 2023-2026 funding phase) will, e.g., add functionality to select required data items from an automatically generated data element catalogue based on the FHIR profiles of the MII CDS, to then automatically extract such data from the DIC FHIR servers and preprocess it for secure data delivery between all data providers, a central data integrating service and the final project coordinator. Further, the contract pipeline is being prepared to support digital signatures. All those steps need to be transparent to researchers and DIC staff, encapsulated within FDPG components and workflows.

References

[1] Gehring S, Eulenfeld R. German Medical Informatics Initiative: Unlocking Data for Research and Health Care. Methods Inf Med. 2018;57(S 01):e46-e49.

[2] Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data Integration for Future Medicine (DIFUTURE). Methods Inf Med. 2018 Jul;57(S 01):e57-e65.

[3] Haarbrandt B, Schreiweis B, Rey S, et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods Inf Med. 2018;57(S 01):e66-e81.

[4] Prokosch HU, Acker T, Bernarding J, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. Methods Inf Med. 2018;57(S 01):e82-e91.

[5] Winter A, Stäubert S, Ammon D, et al. Smart Medical Information Technology for Healthcare (SMITH). Methods Inf Med. 2018;57(S 01):e92-e105. doi:10.3414/ME18-02-0004

[6] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018;57(S 01):e50-e56. doi:10.3414/ME18-03-0003.

[7] https://www.medizininformatik-initiative.de/en/start [last visited on December 4th 2022].

[8] Prokosch HU, Bahls T, Bialke M, et al. The COVID-19 Data Exchange Platform of the German University Medicine. Stud Health Technol Inform. 2022;294:674-678.

[9] Prokosch HU, Baber R, Bollmann P, Gebhardt M, Gruendner J, Hummel M. Aligning Biobanks and Data Integration Centers Efficiently (ABIDE_MI). Stud Health Technol Inform. 2022;292:37-42.

[10] Zenker S, Strech D, Ihrig K, et al. Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: Towards a new German national standard. J Biomed Inform. 2022;131:104096.

[11] https://www.medizininformatik-initiative.de/de/uebersicht-ueber-versionen-der-kerndatensatz-module [last visited on February 22, 2023].

[12] Gruendner J, Deppenwiese N, Folz M, et al. The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. JMIR Med Inform. 2022;10(5):e36709.

[13] Rosenau L, Majeed RW, Ingenerf J, et al. Generation of a Fast Healthcare Interoperability Resources (FHIR)-based Ontology for Federated Feasibility Queries in the Context of COVID-19: Feasibility Study. JMIR Med Inform. 2022;10(4):e35789.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230125

A Model for Multi-Institutional Clinical Data Repository

Karthik NATARAJAN [a,1], Chunhua WENG [a] and Soumitra SENGUPTA [a]
[a] *Columbia University, New York, NY, USA*
ORCiD ID: Karthik Natarajan https://orcid.org/0000-0002-9066-9431, Chunhua Weng https://orcid.org/0000-0002-9624-0214

Abstract. Creating a sustainable model for clinical data infrastructure requires the inclusion of key stakeholders, harmonization of their needs and constraints, integration with data governance considerations, conforming to FAIR principles while maintaining data safety and data quality, and maintaining financial health for contributing organizations and partners. This paper reflects on Columbia University's 30+ years of experiences in designing and developing clinical data infrastructure that synergizes both patient care and clinical research missions. We define the desiderata for a sustainable model and make recommendations of best practices to achieve a sustainable model.

Keywords. clinical data warehouse, data governance, data modeling

1. Introduction

Clinical data is essential to the learning health system. Since the 1980s, Columbia University's Department of Biomedical Informatics (DBMI) has been maintaining a clinical data repository [1], a homegrown web application connected to the repository for clinical care purposes, and a clinical data warehouse that has grown to include the longitudinal electronic health records (EHR) of >6 million patients to support institution wide clinical and translational research. Over the past 35+ years, this sustainable clinical data infrastructure has evolved and witnessed major changes in our organization: i.e., healthcare partners and facilities have grown through mergers and acquisitions, clinical systems have come and gone, and, more recently, our institution has converged on a single EHR system with our partner organizations. Currently, the care environment spans Columbia University Irving Medical Center (CUIMC), Weill-Cornell Medical Center, and NewYork-Presbyterian Hospital in New York City (and vicinity) in a tri-institutional Organized Health Care Arrangement with a single instance of EpicCare EHR for 11 hospitals, 2 faculty practice organizations, and a hospital medical group. In establishing a unified EHR, the leaders of the three institutions have set a clear direction for all, which is that the clinical data is to be equitably used for the success of care, research, finance, and education. This paper reflects on our key considerations and approaches to develop a sustainable data infrastructure and to achieve responsible and balanced tri-institutional management of clinical data for operations and research uses.

[1]Corresponding Author: kn2174@cumc.columbia.edu

2. Current Data Ecosystem

To overcome the challenge of data silos due to diverse data sources, nonintegrated data management systems with heterogeneous schemas, query languages, and APIs, a data lake approach was adopted as a viable solution for providing a schemaless repository for raw data with a common access interface. To date, our clinical data warehouse has been merged into an enterprise data lake consisting of both clinical and non-clinical data from the three institutions, with data feeds mostly in raw forms from EHRs, ancillary systems, and other transactional operational systems using a broad range of methods: HL7 messages, database replication, and batch extracts from source systems. Depending on the type of data, the lake stores data in form of SQL (for clinical data via Microsoft SQL server) and Non-SQL (for Telemetry and device data, via Hadoop from Cloudera) databases. Transformations and quality control checks are applied to the data lake to create curated databases, such as specific data marts (e.g., COVID-19 mart) or generic clinical data warehouses (with data standardized across EHRs using a local terminology system called Medical Entities Dictionary (MED) [2]), or operational datasets for dashboarding and reporting. In the past 5 years, Columbia has participated in or led national data networks such as OHDSI, PCORNet, National COVID Cohort Collaborative, and All of Us Research Program, requiring investment in maintaining heterogeneous common data models. With experience in harmonizing data from over 100 legacy systems, our enterprise data lake is compatible with these diverse data models. The terminology system and provenance mechanisms within the MED not only reconcile disparate data but also serves as a rich information source to understand the richness and limitations of the data. A key observation is that a data ecosystem is and will always be dynamic so that new sources or ETL (Extract, Transform, and Load) outputs should be incorporated constantly with system changes under a common governance.

3. Common Governance

Three institutions have collaboratively established two committees to govern data: the Committee creates policies such as a Data Sharing Agreement (currently 3rd generation in 10+ years) across all institutions, applies the policy towards the requirements for research, quality and operational requests, and sets rules about how access to data is provided and to whom while balancing the security requirements for data requests. For efficiency, routine operational (includes quality) requests are fast tracked, but research and cross-institutional requests are evaluated individually, examining IRB approvals, data use agreements or contracts related to external data sharing, and appropriate scoping. The committees are filled by multidisciplinary key stakeholders such as CIOs, CMIOs, research administration personnel, informatics leaders, leading clinician scientists, finance leaders, analytics leaders, researcher representatives, and clinical data engineers, from all three institutions. Incorporating researchers as stakeholders guarantees data access for responsible research personnel and creates educational opportunities for researchers to learn about the data and to generate more precise and efficient requests. Clinician scientists serve as effective thought leaders and influencers to emphasize to financial leadership that advanced research effectively contributes to exemplary clinical service.

4. Data Access: Cost and Expertise

The sustainability of a clinical data infrastructure and its services cannot be achieved without a fair cost model and a supportive mechanism for training and knowledge sharing. The cost for managing the enterprise data lake infrastructure is funded by tri-institutional operations groups, while data extraction is jointly funded by both operational analytics and research groups. We have previously reported the complexity of clinical data queries [3] and the iterative, human-centered nature of query clarification processes. We also found that use of self-service tool varies by experience and knowledge of users, which can potentially exacerbate the equity of data access. Therefore, we have explored data-driven methods for identifying common data elements needed by researchers [4], but mostly prioritized our effort towards manual service to aid researchers during data access. When a research request is approved, a set of data analysts, called Data Navigators (DN), extract data. DNs develop expertise about the data over time, and several technical approaches are used to disseminate such knowledge to support peer-based learning among DNs. The group of DNs conduct regular webinars on specific topics within the enterprise data lake. A Microsoft Teams group exist for DNs to communicate and exchange lessons learned on specific data. In addition, documentation exists in many different forms – wiki, data catalog, and repository of ETL code. DNs also help close the feedback loop by identifying data errors when they fulfill requests, assisting in making data more complete, and developing code and logic to query certain types of data that is shared across all DNs as well as operational analysts. Turnaround time is a key evaluation metric for DNs. The benefit of the DN model is that costs are shared across the spectrum of the institution, the department or the division, and the individual researcher based on what fits the need most. Each department can employ a data navigator to whom all data requests from that department are directed. Alternatively, a department can work with institutional IT to fund a DN partially and annually, or there can be a fee-for-service model for an individual researcher .

5. Desiderata for Sustainability

Our experience shows that sustaining the success of a clinical research infrastructure, specifically facilitating efficient access to clinical data for both operational and research uses, requires continual demonstration of the value of data and building trust in people and processes through transparency, fairness, partnership, and accountability, as further specified as the following desiderata:

5.1. Strong Partnership Among Stakeholders

All stakeholders must believe that disciplined and timely data availability is key for data driven insight for efficient health care operations, improvement of care quality, innovative clinical and informatics-based research, optimization of health finances, and overall success of the enterprise. All stakeholders must be committed to availability of data and responsible and secure use of data.

5.2. Extensible FAIR Data Infrastructure

The data infrastructure should follow the FAIR (Findable, Accessible, Interoperable, and Reusable) principle for data management. The data infrastructure has to accommodate

the establishment and cataloging of a myriad of data from different vendors and disparate databases with heterogeneous data models. Data ingestion should support diverse methods ranging from file transfer, remote queries, database replication, to incorporating HL7 (and now FHIR) based data feeds. Extensibility entails the ability to easily create different types of curated data marts for subsequent uses.

5.3. Comprehensive Data Governance

The platform and facilities to conduct both should be the same. The governance that establishes policy on how data is to be accessed and distributed must be a collaboration between operations and clinical administration to address the needs of both groups. The common governance is a trust-building activity, where operational, regulatory, and research interests are represented, and shared goals and results are emphasized.

5.4. A Cost-Sharing Model

Fair sharing of the costs of data consolidation, curation, extraction and analysis among the institution, the department, and the researcher is critical to sustain a collaborative research infrastructure. It is also an outcome of the trust models built under the common governance that includes appropriate representation of all stakeholders. Transparency in terms of use and discipline about how data requests are validated guide adherence to the institutional policies, benefiting all stakeholders.

5.5. Evidence of Value Add and Return on Investment

The request intake and fulfillment must be measured and reported to all stakeholders to monitor the efficiency of the system and individuals and identify areas for improvements as needed. Tracking research requests and connecting the requests to subsequent publications or grant awards is a concrete measure of return on investment. In addition, it is important to be able to measure success using survey techniques in collaboration with research administration. Transparency requires that each stakeholder is made aware of the metrics related to their investments.

5.6. Education and Technology Support

Comprehensive documentation and training are necessary in all aspects of the data infrastructure. A key requirement for success is development of personnel such as data engineers, analysts and scientists who understand the depth, nuances, and limitations of data. This is achieved by creating a combined educated workforce that collaborates and educates each other of new data resources.

5.7. Closed Feedback Loop and Support for Knowledge Sharing

It is desirable to close the feedback loop by engaging active contributions of different stakeholders, especially researchers and data engineers, who can report quality problems (or offer correction solutions) as they use the data, in collaboration with clinicians. A forum is needed to enable clinical data knowledge sharing among data users and other stakeholders.

6. Recommendations

In response to the aforementioned desiderata, we arrived at the following recommendations of best practices to improve the sustainability of clinical data infrastructure.

1. Inspire leadership to appreciate the full potential of clinical data for operations and research with latter informing how to improve content and processes.
2. Create an inclusive governance structure of all stakeholders that balances operational needs, security and privacy, and research needs without impeding progress.
3. Develop models of access but with appropriate controls for accountability and monitoring. Ensure that the controls are neither prohibitive nor lax and are set by a governance committee.
4. Identify and implement flexible cost sharing models that are reasonable based on the abilities of multi-level entities: e.g., institution, department, and researcher.
5. Track return of investment and value added and share this information with stakeholders.
6. Create solutions to educate the research community of the process, provide transparency of the process, and demonstrate accountability through metrics.
7. Develop talent with transferable knowledge to maintain continuity of services and provide tools for knowledge capture and exchange. Solicit active feedback from researchers and navigators, empowering them to play an active role in improving data quality.

7. Conclusion

Institutional leadership is critical for successful data infrastructure and effective analytics and research use of data. A collaborative, transparent model encourages proportional cost sharing and development of appropriate data expertise. Shared governance results in responsible sharing of data for secondary use, and in return, data quality is improved by continuous feedback from users. Since volume, variety, velocity, and veracity of health care data will only increase in the future, the infrastructure and use of data has to continually evolve. Strong governance, data engineering, and skilled data personnel are critical for continued success for future infrastructure developments, such as cloud computing, to support both research and operations.

References

[1] Johnson SB. Generic data modeling for clinical repositories [Journal Article]. J Am Med Inform Assoc. 1996;3(5):328-39. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8880680.

[2] Cimino JJ, Johnson SB, Hripcsak G, Hill CL, Clayton PD. Managing vocabulary for a centralized clinical system [Journal Article]. Medinfo. 1995;8 Pt 1:117-20. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8591133.

[3] Hruby GW, Rasmussen LV, Hanauer D, Patel VL, Cimino JJ, Weng C. A multi-site cognitive task analysis for biomedical query mediation [Journal Article]. Int J Med Inform. 2016;93:74-84. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27435950.

[4] Hruby GW, Hoxha J, Ravichandran PC, Mendonca EA, Hanauer DA, Weng C. A data-driven concept schema for defining clinical research data needs [Journal Article]. Int J Med Inform. 2016;91:1-9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27185504.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230126

Enhancing Data Protection via Auditable Informational Separation of Powers Between Workflow Engine Based Agents: Conceptualization, Implementation, and First Cross-Institutional Experiences

Felix ERDFELDER[a,b,c], Henning BEGERAU[a,b], David MEYERS[a,b], Klaus-Jürgen QUAST[a,b], Daniel SCHUMACHER[a,*], Tobias BRIEDEN[d], Roland IHLE[d], Danny AMMON[e], Henner M. KRUSE[e], and Sven ZENKER[a,b,c,1]
[a] *Staff Unit for Medical & Scientific Technology Development & Coordination (MWTek), Commercial Directorate; University Hospital Bonn, Germany*
[b] *Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Germany*
[c] *Department of Anesthesiology and Intensive Care Medicine, University Hospital Bonn, Germany*
[d] *Data Integration Center, Central IT Department, University Hospital Essen, Germany*
[e] *Data Integration Center, IT Department, Jena University Hospital, Germany*
**Present address: Skillbyte GmbH, Cologne, Germany*
ORCiD ID: Felix Erdfelder 0000-0002-4797-1889, Sven Zenker 0000-0003-0774-0725, Henning Begerau 0000-0001-5503-138X, Tobias Brieden 0000-0001-5512-7015, Roland Ihle 0000-0002-8502-7303, Danny Ammon 0000-0001-8960-7316, Henner M. Kruse 0000-0003-0286-8682

Abstract. German best practice standards for secondary use of patient data require pseudonymization and informational separation of powers assuring that identifying data (IDAT), pseudonyms (PSN), and medical data (MDAT) are never simultaneously knowable by any party involved in data provisioning and use. We describe a solution meeting these requirements based on the dynamic interaction of three software agents: the clinical domain agent (CDA), which processes IDAT and MDAT, the trusted third party agent (TTA), which processes IDAT and PSN, and the research domain agent (RDA), which processes PSN and MDAT and delivers pseudonymized datasets. CDA and RDA implement a distributed workflow by employing an off-the-shelf workflow engine. TTA wraps the gPAS framework for pseudonym generation and persistence. All agent interactions are implemented via secured REST-APIs. Rollout to three university hospitals was seamless. The workflow engine allowed meeting various overarching requirements, including auditability of data transfer and pseudonymization, with minimal additional implementation effort. Using a distributed agent architecture based on workflow engine technology thus proved to be an efficient way to meet technical and organizational requirements for provisioning patient data for research purposes in a data protection compliant way.

[1] Corresponding Author: Sven Zenker, Venusberg-Campus 1 53127 Bonn, E-mail: zenker@uni-bonn.de.

Keywords: FHIR, pseudonymization, interoperability, data protection

1. Introduction

The Medical Informatics Initiative (MII) in Germany [1], initiated and funded by the Federal Ministry of Education and Research, has, since 2016, pursued facilitating secondary use of patient data in German University Hospitals by developing a generic, interoperable infrastructure to enable local and cross-site data sharing and usage. In the initial implementation phase of the MII, the SMITH consortium, currently comprising ten German university hospitals, collaborated in realizing and evaluating one of four competitive implementation concepts to achieve this goal [2]. The initial plans to realize these goals by employing an IHE-based infrastructure [2] was found to require adjustment to achieve these goals across the SMITH sites since the complex requirements of secondary use of highly confidential, heterogeneous patient data for research were not mappable to IHE standard compliant processes and interfaces with the available technologies provided by the SMITH commercial partners without breaking standard compliance and thus interoperability with the available resources. In particular, the need to technically and organizationally segregate the information technological (IT) infrastructure at the SMITH sites into two domains – the clinical domain (CD) and the research domain (RD) – was identified. The CD allows processing of all patient related data – including their identifying information – for the purpose of supporting and optimizing patient care, with the legal basis provided by the contractual relationship between patient and healthcare provider, whereas the RD only processes data of patients who have given broad consent [3,4] to secondary use of their pseudonymized patient data for research purposes (Fig. 1). Current German national best practice standards [5] require an informational separation of powers, where no party involved in the pseudonymization process required when performing selective data transfer from CD to RD may at any time simultaneously have access to directly identifying information (IDAT, e.g. a patient's name or date of birth), medical data (MDAT, the "payload"), and definitive pseudonyms (PSN), with the intention to organizationally and technically minimize the risk of malicious or unintentional de-pseudonymization (Fig. 1). To achieve this, the relations between IDAT and PSN are typically entrusted to a trusted third party service (TTA), effectively a third independent domain, which alone can de-pseudonymize by retrieving IDAT for a given PSN, but never sees any MDAT.

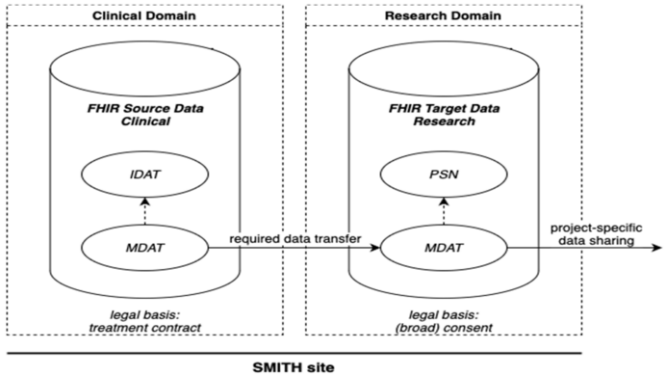


Figure 1. Domain separation and data storage at SMITH sites.

2. Methods

To achieve this goal, the target overall data selection, pseudonymization, and physical data transfer workflow was mapped to an agent-based architecture where a Clinical Domain Agent (CDA) and a Research Domain Agent (RDA) interact with a trusted third party agent (TTA) to jointly execute a distributed workflow. Informational separation of powers is realized via generation of a temporary transport PSN (TPSN) for each potentially identifying data item. This is used to tag MDAT for transferral from CDA to RDA, resolved via TTA by RDA, and deleted irreversibly after successful completion of the entire transferral workflow (Fig. 2). CDA consumes FHIR data from heterogeneous sources while RDA synchronizes to a standard compliant FHIR server. All agents were implemented in Java™ SE 11 using the Spring™ framework [6], allowing lightweight, low-maintenance operation on independent virtual machines in separated network domains. CDA and RDA use the open source version of the Camunda™ workflow engine [7] to auditably execute BPMN workflows modeled using the Camunda™ modeler, while TTA uses gPAS from the MOSAIC toolbox [8] for pseudonym generation and persistence. IDAT and MDAT were represented in MII core dataset FHIR profile [9] compliant JSON. All agent interactions were realized via certificate authentication based, TLS secured REST APIs using Spring Boot and Spring Security mechanisms, including the workflow handover between separated Camunda™ instances via the Camunda™ REST API. IDAT->PSN substitution was extended beyond classical IDAT items to FHIR identifiers that could facilitate identification attacks using a generically parameterizable substitution/deidentification mechanism that made use of the inherent structure of FHIR based data representations.

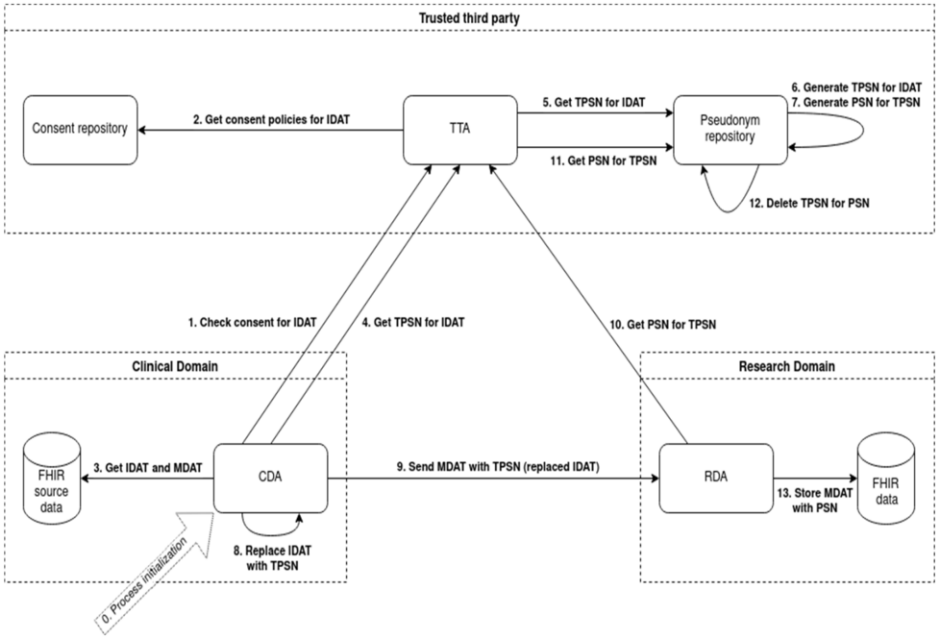


Figure 2. Distributed workflow of CDA and RDA with TTA.

3. Results

Combined usage of the Spring framework and the Camunda™ engine allowed for rapid prototype implementation of the overall system at one SMITH site (Bonn), while a second SMITH site (Essen) implemented the generic deidentification mechanisms. The Camunda™ user interface and audit features, which facilitated straightforward monitoring of transfer processes and rapid post hoc analysis of system behavior, combined with the broad set of convenience features provided by the Spring framework, enabled rapid prototype implementation and verification, with less than three months from initial conceptualization to roll-out to the first non-development SMITH site. Site-specific customization requirements driven by the inter-site infrastructural heterogeneity resulting from very different clinical IT setups at SMITH sites were subsequently identified, extracted, and exposed via Spring declarative configuration mechanisms, allowing SMITH sites with limited software development resources to adapt the system to their specific requirements and thus achieve successful integration. Within 5 months, 3 SMITH sites were thus enabled to demonstrate successful, data protection compliant and fully auditable CD to RD data transfers involving a TTA, contributing to all SMITH sites successfully passing the obligatory MII project audit in spring 2021.

In practice, system stability and performance, even without any investment in specific tuning, was found to be quite satisfactory, with a typical dataset of roughly 5 million FHIR-resources transferred in 5:40 hours.

4. Discussion

The clean decomposition of functionalities and processing steps enforced by the workflow-centric architecture facilitated distribution of development tasks between the Bonn and Essen teams, enabling an agile distributed development workflow without exacerbating integration efforts when merging contributions from the participating development sites. The capabilities of the workflow engine proved crucial to successfully verifying and validating the system within the available timeframe and resource constraints and remain essential to comply with auditability and monitoring requirements imposed by local data protection and IT security regulations. Declarative configuration enabled by the Spring mechanisms proved essential to enable partner sites without sufficient software development capabilities to rapidly adapt the developed system to their local requirements.

Since completion of the primary prototype implementation, the solution has continuously evolved, improving robustness and adding features such as a configurable cohort selection for scheduled data transmission tasks. During this agile and iterative improvement process, the workflow engine based approach proved valuable in reducing the complexity of adaptation tasks and simplifying debugging of system behavior.

Future work will include adaptation of the developed system to the evolving requirements of MII use cases, including managing data types such as biosignals and images not amenable to efficient representation in FHIR, and integration of the developed system with the evolving and increasingly convergent national MII infrastructure of the current funding phase recently initiated, which is currently being constructed from components derived from the results of the four consortia from the previous funding phase.

5. Conclusions

We have shown that an agent-based decomposition of a de-identification and data transfer workflow between different domains of a medical IT architecture involving a trusted third party service to achieve informational separation of powers can successfully, rapidly, and efficiently be implemented and rolled out using an off-the-shelf workflow engine combined with modern Java™ frameworks. Modern workflow engines additionally resolve various overarching concerns of critical relevance when processing sensitive patient data, including full auditability of all data processing steps.

Acknowledgements: The authors wish to thank Dr. Jan Gewehr & team at the SMITH partner site Hamburg Eppendorf for their invaluable contributions to devising the described state of the SMITH architecture and Dr. Thomas Wendt, Sebastian Stäubert & the entire SMITH project management team at the SMITH coordinating site Leipzig for organizational and documentation support. This work was supported, in part, by German Federal Ministry of Education and Research (BMBF) grants no. 01ZZ1803Q, 01ZZ1803P, and 01ZZ1803C.

6. References

[1] Semler SC, Wissing F, Heyder R, German Medical Informatics Initiative. Methods Inf. Med. 2018;57:e50-6. doi:10.3414/ME18-03-0003.

[2] Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, Daumke P, Decker S, Funkat G, Gewehr JE, de Greiff A. Smart medical information technology for healthcare (SMITH). Methods of information in medicine. 2018 May;57(S 01):e92-105. doi:10.3414/ME18-02-0004.

[3] Bild R, Bialke M, Buckow K, Ganslandt T, Ihrig K, Jahns R, Merzweiler A, Roschka S, Schreiweis B, Stäubert S, Zenker S. Towards a comprehensive and interoperable representation of consent-based data usage permissions in the German medical informatics initiative. BMC Medical Informatics and Decision Making. 2020 Dec;20:1-9. doi:10.1186/s12911-020-01138-6.

[4] Zenker S, Strech D, Ihrig K, Jahns R, Müller G, Schickhardt C, Schmidt G, Speer R, Winkler E, von Kielmansegg SG, Drepper J. Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio) medical research: towards a new German national standard. Journal of Biomedical Informatics. 2022 Jul 1;131:104096. doi:10.1016/j.jbi.2022.104096.

[5] Pommerening K, Drepper J, Helbing K, Ganslandt T. Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0. MWV Medizinisch Wissenschaftliche Verlagsgesellschaft; 2014. doi:10.32745/9783954662951.

[6] Johnson R. Expert One-on-One J2EE Design and Development, John Wiley & Sons, 2004.

[7] The Universal Process Orchestrator, *Camunda*. (n.d.). https://camunda.com/ (accessed January 13, 2023).

[8] Bialke M, Bahls T, Havemann C, Piegsa J, Weitmann K, Wegner T, Hoffmann W. MOSAIC–A modular approach to data management in epidemiological studies. Methods of information in medicine. 2015 Jul;54(04):364-71. doi:10.3414/ME14-01-0133.

[9] Basic modules of the MII core data set | Medical Informatics Initiative, (n.d.). https://www.medizininformatik-initiative.de/en/basic-modules-mii-core-data-set (accessed January 13, 2023).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230127

Data Quality Assessment for Observational Medical Outcomes Partnership Common Data Model of Multi-Center

Seol Whan OH[a,b], Soo Jeong KO[a,b], Yun Seon IM[a,b], Surin JUNG[a], Bo Yeon CHOI[a,b], Jae Yoon KIM [a,b], Wona CHOI [a,1] and In Young CHOI [a,2]
[a] *Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea*
[b] *Department of Biomedicine &Health Sciences, The Catholic University of Korea, Seoul, Republic of Korea*
ORCiD ID: Seol Whan OH https://orcid.org/0000-0002-0328-9634
Soo Jeong KO https://orcid.org/0000-0002-6550-9188
Yun Seon IM https://orcid.org/0000-0002-2510-1380
Surin JUNG https://orcid.org/0000-0002-3314-3185
Bo Yeon CHOI https://orcid.org/0000-0003-1311-2645
Jae Yoon KIM https://orcid.org/0000-0001-8847-9586
Wona CHOI https://orcid.org/0000-0003-0269-6374
In Young CHOI https://orcid.org/0000-0002-2860-9411

Abstract. The amount of research on the gathering and handling of healthcare data keeps growing. To support multi-center research, numerous institutions have sought to create a common data model (CDM). However, data quality issues continue to be a major obstacle in the development of CDM. To address these limitations, a data quality assessment system was created based on the representative data model OMOP CDM v5.3.1. Additionally, 2,433 advanced evaluation rules were created and incorporated into the system by mapping the rules of existing OMOP CDM quality assessment systems. The data quality of six hospitals was verified using the developed system and an overall error rate of 0.197% was confirmed. Finally, we proposed a plan for high-quality data generation and the evaluation of multi-center CDM quality.

Keywords. Data Quality, Common Data Model, Data Quality Management System

1. Introduction

Due to the increasing importance of healthcare data, there has been a growing interest in the study of data collection and management in recent years [1,2]. However, the structure of healthcare data is different for each hospital, making it difficult to conduct multi-institutional research related to data collection [3]. Therefore, many institutions have

[1] Corresponding Author: Wona CHOI, E-mail: choiwona@gmail.com
[2] Corresponding Author: In Young CHOI, E-mail: iychoi@catholic.ac.kr

established a common data model (CDM), thus laying the groundwork for cooperative methodological or clinical innovation [3].

Data quality is one of the main issues when gathering and managing large amounts of data [1,4]. Data quality issues arising from the nature of the source data and deficiencies in the data conversion process itself can affect the actual usefulness and reliability of CDM data [1,4]. Therefore, several research institutes including the Observational Health Data Sciences and Informatics (OHDSI) have developed and distributed data quality tools for CDM. However, the implementation of this tool requires considerable database knowledge and the applied verification rules cannot be easily changed [5].

To address the aforementioned limitations, this study sought to create a more user-friendly CDM data quality assessment system for multi-center quality evaluation. Additionally, advanced evaluation rules were developed using data quality evaluation rules created by several research institutes. This will allow for the evaluation of the quality of the data derived from several institutions that have formed a CDM, as well as the identification of the quality of the CDM establishment for each institution. More importantly, the proposed tool provides a novel means for high-quality data collection, thus overcoming current bottlenecks in data acquisition and management in the health sector.

2. Methods

This study created a system for multi-center CDM data quality assessment, as well as advanced evaluation rules for quality assessment. The CDM data quality assessment system, which was loaded with advanced evaluation rules, was then used to conduct multi-center CDM data quality evaluation.

2.1. Development of CDM Data Quality Assessment System

The system was developed based on the Observational Medical Outcomes Partnership CDM (OMOP CDM), a widely known and representative CDM. The system was developed by accessing the hospital's CDM database and the results were extracted through the loaded evaluation rule query. All database queries were executed using PostgreSQL. Figure 1 illustrates the architecture of the CDM data quality assessment system developed herein.

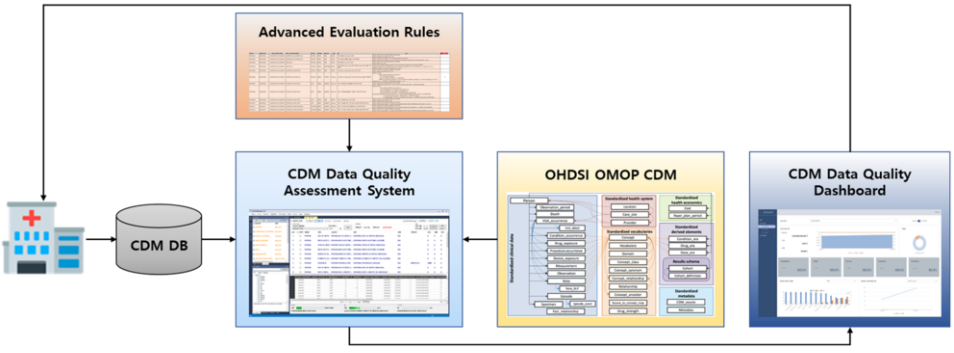


Figure 1. CDM Data Quality Assessment System Architecture.

2.2. Advanced Data Quality Evaluation Rules

The evaluation rules loaded into the system were based on DQ4HEALTH, a multi-institutional medical data quality measurement model [1]. The data quality rules were developed using five quality dimensions.

- Completeness: This rule evaluates whether null values are entered among the required items that must contain values for each column according to the table definition, such as the person number in the person table [1,2,6].
- Uniqueness: The primary key (pk) column, which is an identification value in the database, must not have duplicate values. This rule is particularly important to ensure that unique identifiers such as those in the condition occurrence ID column of the condition occurrence table are not duplicated [1,7].
- Validity: Data for dates or measurement values must be within a valid range of values. Given that the data format is defined according to the table definition, there is a rule to check whether the data fits the format properly. For example, the number of specimens should be greater than zero and birth month must be one or two digits [1,8].
- Consistency: Some values are entered by referencing between variables in other columns or tables. This rule evaluates whether the values referenced from other tables are appropriate. For example, a drug concept ID must be a drug domain [1,8].
- Accuracy: This rule verifies whether the expression value of an object is accurately reflected. Specifically, the rule checks whether a value calculated from multiple values is correct or whether dates are in chronological order. For example, the measurement date must be between the date of birth and the date of death [1,8,9].

To optimize the evaluation rules, we compared them with other data quality rules. IQVIA, a healthcare research and service company, is supporting the construction of OMOP CDM, and the data quality rules used in this CDM were evaluated in the present study [10]. Additionally, our study also assessed the Data Quality Dashboard (DQD) evaluation rules provided by OHDSI [6].

Evaluation rules were developed by applying the DQ4HEALTH model to OMOP CDM v5.3.1, after which mapping analysis was performed with the provided IQVIA and DQD evaluation rules to obtain a set of more advanced evaluation rules (Figure 2).

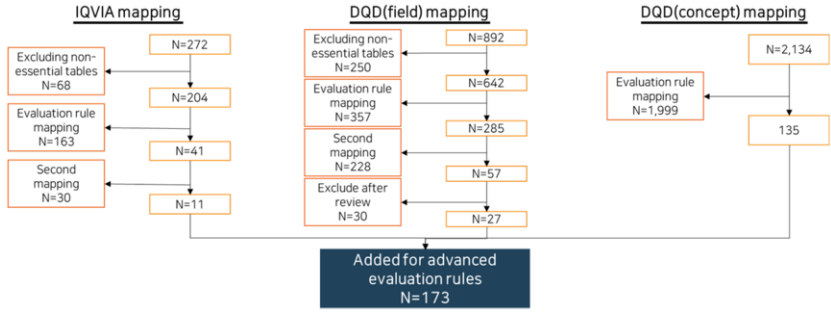


Figure 2. Data Quality Evaluation Rules Mapping Flowchart

2.3. Verification of Multi-Center Data Quality using The System

The developed system was used to evaluate the quality of OMOP CDM data from six hospitals. Considering the development environment of each hospital, online and offline installation versions were separately developed and distributed. After quality evaluation, the results were reviewed to assess the quality of the CDM data from each hospital.

3. Results and Discussions

We developed a CDM data quality evaluation system loaded with verification rule queries to evaluate the quality of multi-institutional OMOP CDM data (OMOP CDM v5.3.1). A total of 179 verification rules were developed by applying all five quality dimensions of the DQ4HEALTH model to 13 tables selected as essential tables at the time of construction.

Table 1. The Number of Advanced Verification Rules

Table	The Number of Verification Rules	Totals
PERSON	28	
OBSERVATION_PERIOD	14	
SPECIMEN	16	
DEATH	11	
VISIT_OCCURRENCE	27	
VISIT_DETAIL	32	
PROCEDURE_OCCURRENCE	101	2,433
DRUG_EXPOSURE	33	
DEVICE_EXPOSURE	26	
CONDITION_OCCURRENCE	229	
MEASUREMENT	1,852	
NOTE	22	
OBSERVATION	42	

A total of 2,433 CDM data quality verification rules were optimized and loaded into the CDM data quality evaluation system (Table 1). Additionally, server management, verification execution and verification report functions were incorporated into the system so that anyone can verify CDM data quality. The data quality verification system was specifically designed to be user-friendly.

CDM data quality verification was performed by six organizations (A, B, C, D, E, F) that had built OMOP CDM v5.3.1 using a system loaded with advanced verification rules. Quality verification was conducted based on the CDM updated by each of the six hospitals as of September 2022, and the error rates were 0.208%, 0.217%, 0.218%, 0.112%, 0.228%, and 0.181%, respectively. The combined error rate of all hospitals was only 0.197%. Our results thus confirmed that the error rates were very low (Table 2).

Table 2. Evaluation results of 6 hospitals based on the CDM data quality verification system

Hospitals	The Number of Patients	The Error Rate
A	927,997	0.207631%
B	1,951,727	0.216930%
C	1,005,002	0.218432%
D	866,168	0.112487%
E	424,752	0.228015%
F	1,159,941	0.181156%

4. Conclusions

Here, we developed a CDM data quality assessment system for OMOP CDM, which is being built and populated by many hospitals. IQVIA and DQD evaluation rules were analyzed to develop advanced quality rules, and these were loaded into the CDM data quality assessment system to allow hospitals to easily assess high-level data quality. For demonstration purposes, data quality evaluation was performed for six hospitals using this system, and the extracted results were checked to present the current status and direction of CDM quality for each hospital.

One of the major constraints of this study is that the number of institutions subject to CDM quality evaluation was limited. All six hospitals verified were generating high-quality CDM data, which was confirmed by our study results. Starting with this study, the system has been distributed to a larger number of target institutions, and CDM data quality evaluation is in progress to assess various datasets from each hospital. Additionally, although the data quality for each hospital was evaluated, a comparative study must also be conducted to evaluate data quality after the implementation of an improvement plan.

Despite the limitations of this study, we developed a CDM data quality assessment system that does not require much database knowledge. This will enable each institution to perform CDM quality verification more easily. Additionally, a dashboard was developed to assess and compare the quality of the data from each institution. These tools will facilitate the evaluation of data quality when building OMOP CDMs. In turn, this will contribute to securing high-quality CDM data, thus enabling each institution to make informed decisions on data improvement.

References

[1] Kim KH, Choi W, Ko SJ, Chang DJ, Chung YW, Chang SH, et al. Multi-Center Healthcare Data Quality Measurement Model and Assessment Using OMOP CDM. Applied Sciences. 2021 Jan;11(19):9188.

[2] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016 Sep 11;4(1):1244.

[3] Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. J Am Med Inform Assoc. 2021 Sep 18;28(10):2251–7.

[4] Lynch KE, Deppen SA, DuVall SL, Viernes B, Cao A, Park D, et al. Incrementally Transforming Electronic Medical Records into the Observational Medical Outcomes Partnership Common Data Model: A Multidimensional Quality Assurance Approach. Appl Clin Inform. 2019 Oct;10(5):794–803.

[5] Callahan TJ, Bauck AE, Bertoch D, Brown J, Khare R, Ryan PB, et al. A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. EGEMS (Wash DC). 2017 Jun 12;5(1):8.

[6] Huser V, DeFalco FJ, Schuemie M, Ryan PB, Shang N, Velez M, et al. Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets. EGEMS (Wash DC). 2016;4(1):1239.

[7] Amicis F de, Batini C. A methodology for data quality assessment on financial data. 2004 [cited 2023 Mar 5]; Available from: https://www.e-periodica.ch/digbib/view?pid=sco-003:2004:4::559

[8] Wand Y, Wang RY. Anchoring data quality dimensions in ontological foundations. Commun ACM. 1996 Nov;39(11):86–95.

[9] Xiao Y, Bochner AF, Makunike B, Holec M, Xaba S, Tshimanga M, et al. Challenges in data quality: the influence of data quality assessments on data availability and completeness in a voluntary medical male circumcision programme in Zimbabwe. BMJ Open. 2017 Jan 27;7(1):e013562.

[10] Candore G, Hedenmalm K, Slattery J, Cave A, Kurz X, Arlett P. Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children. Clin Pharmacol Ther. 2020 Apr;107(4):915–25.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230128

Timeline of and Expectations for the National Medication List in Sweden

Mikael HOFFMANN[a, b, c, 1]
[a] *The NEPI Foundation, Stockholm, Sweden*
[b] *Department of Health, Medicine and Caring Sciences, Linköping University, Sweden*
[c] *The Swedish Society of Medicine, Sweden*
ORCiD ID: Mikael Hoffmann https://orcid.org/0000-0001-9420-3316

Abstract. It is important to separate the continuous chains of medications orders (or decisions by the prescriber) in modern electronic health records from the one-way communication of prescriptions from healthcare to pharmacies. To support the self-administration of prescribed drugs the patient needs a continuously updated list of medication orders. For the NLL to function as a safe resource for the patient, it is necessary to have the information updated, curated, and documented by prescribers in a one-step process within the electronic health record. Four of the Nordic countries have chosen separate ways trying to achieve this. The experiences and obstacles during the introduction of the mandatory National Medication List (NLL) in Sweden and the resulting delays are described. The planned integration for 2022 is now delayed to 2025 and will probably only be achieved in 2028 or even 2030 in some regions.

Keywords. Medication management, shared information lists, user-centred design.

1. Introduction

An electronically shared medication list (SML) can be defined from different perspectives. Focus can be on the medication orders, i.e. the decisions regarding treatment with drugs that a prescriber (most often a physician) is responsible for, or on the prescriptions, i.e. the one-way communication from the prescriber to the dispensing pharmacist.

The term prescription is often used to describe the decision by the physician (the medication order) as well as the communication of information from a prescriber to the dispensing pharmacist, for instance, as an e-prescription. It is essential to distinguish between the decision and the communication since not all medication orders will be communicated, for instance, when a dose is temporarily changed, or the treatment with a drug is stopped prematurely. In addition, medication orders are valid weeks to months after the date of the last dispensation of a prescription.

If the goal of a SML is to support the patient to self-administer drugs at home and to provide prescribers, other healthcare personnel, and pharmacists with a complete and up-to-date list of a patient's medication orders, then it is not sufficient to gather all valid prescriptions communicated to the pharmacists and all dispensed drugs. The challenge is whether to create a shared list of continuously updated medication orders from

[1] Corresponding Author: Mikael Hoffmann, MD PhD MPH, mikael.hoffmann@liu.se

different prescribers or a shared list of prescriptions and dispensed drugs and how to reconcile them with each other through integration with the electronic health records (EHRs) and work processes capturing all medication orders into the updated SML.

The Nordic countries have taken different approaches [1] primarily focusing on either medication orders or prescriptions when constructing a SML. Denmark [2] was the first of the four countries to establish a SML with a nationwide mandatory platform where a medication order in the electronic health record, EHR, automatically updates the SML. Norway [3] also primarily focuses on supporting physicians and patients with a SML integrated into the EHR consisting of all medication orders, thus achieving updated information on valid prescriptions for pharmacists. Finland [4] and Sweden [5] have focused on gathering all e-prescriptions and dispensations in parallel or in a 2nd step providing tools when integrated into EHRs for prescribers to update the prescriptions. Similar initiatives exist or are under development in other countries with different expressed goals and technical solutions.

In Sweden, the National Medication List (Nationell läkemedelslista, NLL) is a new mandatory register at the E-health Agency of Sweden (E-hälsomyndigheten, eHM) [5]. The EHM is also the authority responsible for the Swedish participation in the eHealth Digital Service Infrastructure, EHDSI, including ePrescription and eDispensation. NLL covers all issued prescriptions for dispensations at pharmacies and the dispensed medications for up to five years. The stated goal of NLL is to *"provide the healthcare, pharmacies and the patient with the same information about the medications prescribed and dispensed to the patient"* i.e. not medication orders [5]. The 21 regions providing most of Sweden's healthcare have EHRs and mandatory e-prescriptions. Thirteen regions will switch to new EHRs between 2023 and 2030, resulting in 17 counties using Cambio Cosmic, two Cerner Millennium and two other solutions.

This paper aims to give an insight into the timeline for and an understanding of identified obstacles to implementing a SML in Sweden.

2. Method

This text is a scoping review of statutes, commission reports, articles and letters in Swedish describing the introduction of the National Medication List in Sweden.

3. Timeline

During 2000–2015 two successive projects, PALL (the patient's medication list) and NOD (the national database of medication orders aiming to collect all medication orders within the healthcare system) were initiated by the regions. Legal issues stymied the development of PALL while NOD was abandoned when the government in 2016 suggested a mandatory list of all prescriptions, later named NLL. The law from 2018 stipulated the start of the new register in May 2021, with the last date for the current e-prescription format in June 2022. This last date was postponed to May 2023 due to the covid-19 pandemic. Due to further delays, the government in late 2022 proposed to postpone the mandatory integration with EHRs to December 2025.

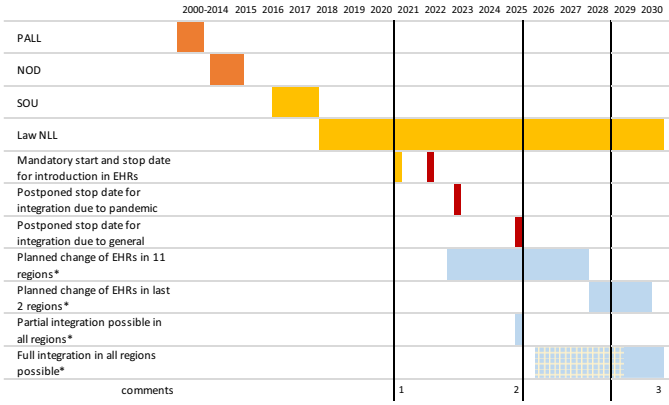


Figure 1. Timeline of the development and deployment of the shared medication list in Sweden. *Estimated plans as interpreted from SALAR's open letter [7]. Numbers – see comments in the text.

3.1. First step – access to information for patients and prescribers through a separate log-in plus publication of specifications for developers of EHRs

With the 1st step of the introduction of NLL after May 2021 (Figure 1, #1), physicians can access and, from late 2021, revise prescriptions. In addition, a new interface and extended history (five years) of dispensed medications are available for both patients (Läkemedelskollen) and prescribers (Förskrivningskollen, intended as a backup system). Läkemedelskollen has two medication lists for the patient, one for prescriptions and one for dispensed medications. The trade name is highlighted in both lists, with the substance name available. Frequent mandatory generic substitution at pharmacies (in 48% of all dispensations 2022) has led to a mismatch between prescribed and dispensed trade names often confusing patients.

3.2. Second step – stepwise integration

HL7 Fast Health Interoperability Resources (FHIR) is used for the planned integration of NLL with EHRs. The integration was supposed to be fully deployed before June 2022 but has been delayed for several years. In the 13 regions planning to switch EHR during the next few years, there was a lack of interest in allocating resources for a fast-track development that would become redundant within a few years. For the eight regions already using Cambio Cosmic, a delay in published specifications and, more importantly, problems in envisioning how a full integration could be devised triggered concerns expressed in an open letter in November 2019 focusing on lack of support for processes in the healthcare [6].

In early 2022 a risk analysis of the process of introducing NLL was performed by the network of Chief Medical Officers responsible for patient safety among regions and large private providers [7]. The risk analysis focused on the need for more time to substitute the current format of e-prescriptions in time through updated EHRs after May 2023. Even more critical from a strategic standpoint was that the specifications developed by eHM focused on the one-way and non-continuous process of prescriptions – as opposed to continuous chains of medication orders both for in-patient and out-patient medication orders in the EHRs. Based on this, a full integration supporting the everyday

work processes in healthcare, avoiding double documentation and additional tasks for the physicians, was deemed impossible to develop and deploy within the time frame available by the network of Chief Medical Officers.

In August 2022, The Swedish Association of Local Authorities and Regions (SALAR) informed the Ministry of Health and Social Affairs that the regions could not implement NLL in the EHRs as mandated by law in May 2023 [8]. The assessments of a possible schedule from SALAR were:

- Access to information in NLL through read-only by users (i.e., not a full integration into the list of medication orders) of EHR in all regions is not possible before 2025 due to the time needed to establish FHIR in established EHRs.
- Writing from EHR to NLL (substituting the e-prescription with the new FHIR-based format) will not be possible for all regions until 2028–2030.

The assessments were also predicated on a list of necessary established solutions and some uncertainties regarding possible consequences of the new EU Medical Device Regulations introduced in 2021. In addition, technical issues regarding the identification and authentication of users were raised.

SALAR described the main problem as the need for significant updates in the EHRs, necessary since the EHRs are based on continuous medication orders and not the logic based on current intermittent prescriptions devised by the E-health Agency for the NLL. Considerable technological and logistical challenges were identified to provide a seamless user experience for healthcare personnel.

The government in December 2022 proposed to postpone the mandatory integration of EHRs with NLL from May 2023 to December 2025 Figure 1, #2).

3.2.1 Integration – Writing from EHR to NLL

A time-limited solution by the E-health Agency for receiving e-prescriptions, translating, and writing these to NLL – *"the Transformator"* – is a critical service to facilitate the continued use of current EHRs. Before integration, it is necessary to avoid both documenting medication orders in the EHR and communicating in parallel in a separate web-based process with the E-health Agency. This complexity is also a focal point of SALAR's analysis and the main reason for presenting 2030 as a probable date for the integration to be available in all regions (Figure 1, #3). Another aspect is that medication orders in the EHRs that do not result in a traditional e-prescription have to update NLL.

3.2.2 Integration – Writing from NLL to EHR

A less well-understood complexity is the need for automated transfer of information about prescribed and dispensed medications from the NLL to the EHR. When information from a list of prescriptions is supposed to update the list of medication orders in another EHR, the necessary information to link this to an existing or introduce a new ordination chain in the EHRs is lacking. The problem also highlights the possible need to introduce new information in the process *"write from EHR to NLL"*, linking the information transferred to NLL to a specific chain of medication orders in the EHR from where it originated, see 3.2.1. These aspects are not covered in the analysis by SALAR, where *"writing from NLL to EHR"* is interpreted not as an integration and thus presented as possible to achieve in 2025.

4. Discussion

The term *"shared medication list"* implies that it is a list of the patient's ordered medication that can support self-administration by the patient directly or with the help of a resource person without medical training. A compilation of valid prescriptions will, however, not fulfil that expectation.

NLL will not by itself significantly improve the quality of the information available for dispensation unless it is integrated in the EHRs and the work processes of prescribers. As highlighted by the risk analysis performed in 2022 [7], a technical compilation of prescriptions made available to the patient daily requires continuous updating and curating of the content by a trained professional to fulfil the demands on a medication list. Correct information will be possible only when the NLL is fully integrated into the EHRs, and a standard work process among prescribers to update and curate NLL through a one-step process in the EHR has been deployed.

5. Conclusions

The introduction of NLL is a positive step, but the perspectives of human, organization and technology [9] still need to be addressed. The need to focus on medication orders to provide the patient with an updated medication list as in Denmark [2] and Norway [3] is relevant for similar projects in Europe if the goal is to support the patient.

Before a full integration in the EHRs has been achieved, most of the benefits of the register cannot be fulfilled. A narrow focus on prescription handling has complicated both integrations in current and new EHRs and in supporting the development of a future common work process by prescribers and pharmacists.

Shared responsibility for finding common ground between authorities, healthcare, pharmacies, prescribers, and pharmacists is needed to achieve the true potential of NLL as a SML for patients as well as for healthcare personnel and pharmacists.

References

[1] Hammar T, Bergmo TS, Bülow C, et al. Nationally shared medication lists – description of systems in the Nordic countries. *Conference Proceeding Medical Informatics 2023 (accepted full paper)*

[2] Fælles Medicinkort (in Danish). Available at: https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-faelles-medicinkort (accessed 4 March 2023)

[3] Program pasientens legemiddelliste (in Norwegian). Available at: https://www.ehelse.no/programmer/program-pasientens-legemiddelliste (accessed 4 March 2023).

[4] Kanta.fi - Citizens. Available at: https://www.kanta.fi/en/ (accessed 4 March 2023).

[5] Nationella läkemedelslistan (in Swedish). Available at: https://www.ehalsomyndigheten.se/yrkesverksam/nationella-lakemedelslistan/ (accessed 10 January 2023).

[6] Efterlängtad lista riskerar bli hatobjekt 2022 (in Swedish). Available at: https://www.dagensmedicin.se/alla-nyheter/nyheter/efterlangtad-lista-riskerar-bli-hatobjekt/ (accessed 10 January 2023).

[7] Riskanalys av Nationella läkemedelslistan 2022. Available at: https://sverigescheflakare.se/images/Chefslakare/Riskanalys-av-Nationella-lkemedelslistan-2022.pdf (accessed 4 March 2023)

[8] Underlag från regionerna om bortre tidsgräns NLL (in Swedish). Available at: https://assets.ctfassets.net/e03pgm1m5c6m/wVvcHxyhpGClusmXf9W0H/7cdc29e18bdd99b3599c5b49a1a81b5e/Underlag_fra__n_regionerna_om_bortre_tidsgra__ns_NLL-20220830.pdf (accessed 4 March 2023).

[9] Yusof MM, Kuljis J, Papazafeiropoulou A, et al. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). *Int J Med Inf* 2008; 77: 386–398.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230129

Enriching Remote Monitoring and Care Platforms with Personalized Recommendations to Enhance Gamification and Coaching

Antonios PARDOS[a,b], Parisis GALLOS[a,b], Andreas MENYCHTAS[a,b]
Christos PANAGOPOULOS[b] and Ilias MAGLOGIANNIS[a,1]
[a] *Computational Biomedicine Research Lab, Department of Digital Systems, University of Piraeus, Greece*
[b] *Bioassist S.A., Greece*
ORCiD ID: Antonios Pardos https://orcid.org/0000-0002-0663-7382
Parisis Gallos https://orcid.org/0000-0002-8630-7200
Andreas Menychtas https://orcid.org/0000-0002-4510-5522
Christos Panagopoulos https://orcid.org/0000-0001-9282-0919
Ilias Maglogiannis https://orcid.org/0000-0003-2860-399X

Abstract. Patients' remote monitoring platforms can be enhanced with intelligent recommendations and gamification functionalities to support their adherence to care plans. The current paper aims to present a methodology for creating personalized recommendations, which can be used to improve patient remote monitoring and care platforms. The current pilot system design is aimed to support patients by providing recommendations for Sleep, Physical Activity, BMI, Blood sugar, Mental Health, Heart Health, and Chronic Obstructive Pulmonary Disease aspects. The users, through the application, can select the types of recommendations they are interested in. Thus, personalized recommendations based on data obtained by the patients' records anticipated to be a valuable and a safe approach for patient coaching. The paper discusses the main technical details and provides some initial results.

Keywords. Gamification, pHealth, recommendations, coaching, eHealth, IT systems, Rule-based Expert System, patient telemonitoring

1. Introduction

Health Information systems often include advanced features aimed to managing diseases and promoting healthy living. [1-3]. Meanwhile, the idea of intelligent patient coaching is also widely accepted in these systems because it can provide users with a more consistent, safer, and more reliable guidance through tailored information and recommendations [4,5]. Additionally, patients' remote monitoring and care can be also enhanced with gamification and functionalities, as recommendations, that support their adherence to care plans [6-8]. The current paper aims to present a methodology for

[1] Corresponding Author: Ilias Maglogiannis, Department of Digital Systems, University of Piraeus, 80, M. Karaoli & A. Dimitriou St., 18534 Piraeus, Greece; E-mail: imaglo@unipi.gr.

creating personalized recommendations and coaching schemes, which can be used in patient remote monitoring and care platforms. The adoption of such technologies may contribute to higher patient engagement on the care plan as well as in a healthier lifestyle and better disease management.

2. Methods

The proposed recommendations' module aims to support patients in achieving a healthier lifestyle. The goal is to "educate" users to develop habits that will improve their daily life quality and their health status. At the same time, another goal is to make sure that these habits adhere to WHO and other relevant organizations' guidelines. [9,10]. The presented initial system is intended to support and educate patients by making recommendations for Sleep, Physical Activity, BMI, Blood sugar, Mental Health, Heart Health, and Chronic Obstructive Pulmonary Disease issues. The system functions as a ruled based "intelligent agent" [11-13], converting the user's data into guidelines for a improving their lifestyle. Domain-specific knowledge is required to model the "environment" where the agent will convert the user's data into recommendations. Furthermore, the need of a scoring mechanism is required in order to assess the user's data with the coordinated recommendations, additionally, the score is used by the *Credit* generator mechanism. A mechanism exploits the user's score to generate the *credits*, a points-like object that is used to enhance the gamification approach [14,15]. Using these *credits*, the user can reclaim some gaming benefits that the application provides. The integration of serious games (mental games and exergames) in the system, facilitates the development of additional Personal Health Record (PHR) data, which is subsequently utilized as an indication to evaluate the user's mental health and mobility. The *Intervention* mechanism inquires for the proper recommendation according to the user's score. This component produces the appropriate domain recommendation. Figure 1 illustrates the aforementioned flow.

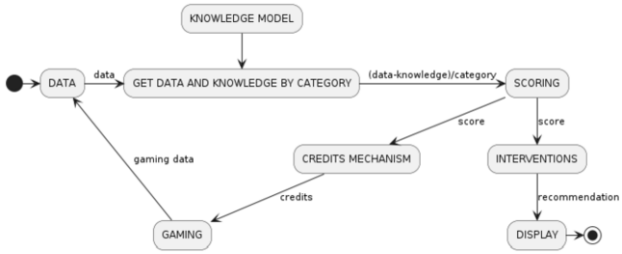


Figure 1. The recommendations flow in the system

2.1. Data Source

Personal Health Records (PHRs) are used as input for the recommendation module. The PHRs are collected by smart devices (mainly smart watches) via third-party applications, by Bluetooth devices or by user's manual input if the user has not any available interconnected devices for measurements like weight or blood-sugar. Furthermore, external data such as weather forecast can be used to make strong-evidence inferences on some recommendations, including those for patients with Chronic Obstructive Pulmonary Disease (COPD), or soft-evidence inferences on other recommendation types like physical activity recommendations.

2.2. Domain-specific knowledge

In Figure 1, the *KNOWLEDGE MODEL* represents the guidelines given by the World Health Organization, other National organizations such as the American Heart Association [9], other medical advisory organizations like Sleep Foundation [10], or individual healthcare physicians. For each domain, the KNOWLEDGE MODEL is encoded as a set of multivariate objects. A default knowledge has been stored in a structure in which each domain contains information about a goal, an acquisition technique, a reference period, and a condition. An example of how such a model could be represented in each domain is shown in Table 1. If needed, physicians can modify goals, periods, and conditions in every domain for each patient.

Table 1. Example of the encoded KNOWLEDGE MODEL

Category	Acquisition	Goal	Units	Reference Period	Condition
Sleep	smart watch	420 - 540	minutes	DAILLY	RANGE
Physical Activity	smart watch	150	minutes	WEEKLY	≥
BMI	weight	18.5 - 25	-	WEEKLY	RANGE
Blood Sugar	Sensor (Blood sugar)	100	mg/dL	DAILLY	≤
Mental Health	Games	2	-	DAILLY	≥
Heart Health	Physical activity, BMI		-	WEEKLY	-
COPD	Weather data, Physical activity	-	-	DAILLY	-

The *KNOWLEDGE MODEL* can generate complicated elements such as *Heart Health* and *COPD*, which are objects that rely on the success of other primitives such as *Physical activity*. The *KNOWLEDGE MODEL* is discussed in more detail in the subsequent section.

2.3. Modeling The Knowledge

The recommendation module is a mechanism which primary function is to encourage and alert the users in developing healthier habits. The *KNOWLEDGE MODEL* consists of rules. The representation of each rule is configured as shown in Figure 2. The category variable has the name of the recommendation domain (physical activity, sleep, BMI, and other). For some categories we only need some aggregation on the user's PHRs data and then to evaluate those data within the KNOWLEDGE MOLEL. *Physical activity*, *Sleep*, and *Blood sugar* are examples of such types. These types of categories are referred as *"Primitive"* types in the *KNOWLEDGE MODEL*. Other types, such as *Heart Health*, *COPD*, dependent on the success of some primitive types. These are called *Mixed* types in our *KNOWLEDGE MODEL*. Figure 2 also depicts instances of primitive and mixed elements where we know that (a) a daily sleep should be between 7 and 9 hours, and (b) for the *Heart health* element, a systematic blood pressure measurement and success in both physical activity and BMI assessments are necessary. The current object explicitly sets that the Primitive objects, PHYSICAL ACTIVITY and BMI should be accomplished. Additionally, it defines that an individual should perform at least 7 blood-pressure measurements in a week. A Physician may change the goals and dependencies to create new tailor-made rules (elements in the KNOWLEDGE BASE) based on the different patient's conditions.

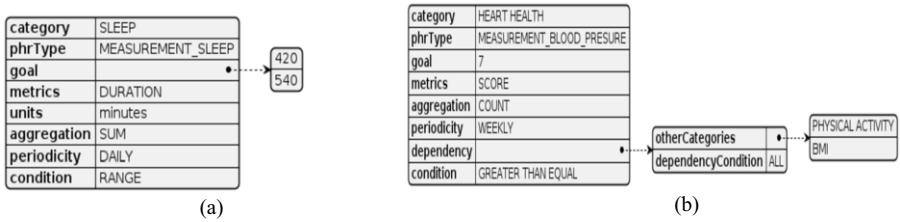


Figure 2. The structure of a Primitive element in the KNOWLEDGE MODEL (a), the structure of a Mixed element where the goal, metrics, aggregation and periodicity are optional (b).

2.4. Scoring - Recommendations - Credits

Based on the above, the state of each individual user is evaluated using their PHRs. For this purpose, firstly *Primitive* elements (such as Physical activity, Sleep) and then *Mixed* elements (such as *Heart health*) information is examined. As a result, a score based on the guidelines (rules) and the user's data is computed and forwarded to the subsequent processes, which involve the "credit construction" and "recommendation creation". Three different levels of scoring are defined. The first level indicates that the goal was met. The second level indicates that a little more effort is required, and the third indicates that significant more effort is required by the patient. A recommendation is formulated and displayed on the user's application screen based on the user's score in each category. Furthermore, the credits earned can be used to play a variety of games through the app.

3. Results and Discussion

The application allows the users to select the types of recommendations according to their profile. Relevant recommendations based on their preferences are displayed on their application screen. Figure 3 illustrates the user settings (a) and the related recommendations (b).

Figure 3. (a) The UI widget from which the user selects the categories. (b) Health recommendations in the user's screen. (c) A snapshot from a recommendation card.

Each suggestion card has a unique color palette, based on the recommended category. The card includes an image relevant to the category, a title, and a message. In addition, a success indicator in the form of a "ring" appears. When the goal is reached, the ring appears to take a 3D shape (Figure 3 b). The user can view his status history for the given recommendation field as a graph by clicking a button on the interface. He can also find more about the source of the recommendation by clicking the information icon. Figure 3 (c) presents the recommendation card and its elements.

4. Conclusions

The proposed recommendations' tool can be applied in patient remote monitoring and care platforms to enhance gamification and the patients' coaching. Personalized recommendations based on data obtained by the patients' records seems to be a valuable and a safe approach for patient coaching. A limitation is that the system's knowledge field displays the key condition dependence factors. In fact, each condition may be dependent on more than the aspects reflected in the created model. To keep the approach simple for both the user and the inference process, we are currently excluded scenarios which have smoking, drinking, meals, nutritional value as factors. The proposed approach assessment by different types of patients is one of the future works of our research.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: MediLudus Personalised home care based on game and gamify elements T 2EKΔK-03049).

References

[1] DiPietro J, et al. Computer- and Robot-Assisted Therapies to Aid Social and Intellectual Functioning of Children with Autism Spectrum Disorder. Medicina (Kaunas). 2019 Aug 5;55(8):440
[2] Demiris G, Thompson H. Smart homes and ambient assisted living applications: from data to knowledge-empowering or overwhelming older adults? Contribution of the IMIA Smart Homes and Ambiant Assisted Living Working Group. Yearb Med Inform. 2011;6:51-7. PMID: 21938325.
[3] Panagopoulos C, et al. Utilizing a Homecare Platform for Remote Monitoring of Patients with Idiopathic Pulmonary Fibrosis. Advances in Experimental Medicine and Biology. 2017;989.
[4] Ammar A, et al. Applying digital technology to promote active and healthy confinement lifestyle during pandemics in the elderly. Biol Sport. 2021 Sep;38(3):391-396.
[5] Pardos A, Menychtas A, Maglogiannis I. Introducing Gamification in eHealth Platforms for Promoting Wellbeing. Stud Health Technol Inform. 2022 Jan 14;289:337-340
[6] Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform. 2017 Jul;71:31-48. doi: 10.1016/j.jbi.2017.05.011. Epub 2017 May 20. PMID: 28536062.
[7] Miller AS, Cafazzo JA, Seto E. A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics J. 2016 Jun;22(2):184-93.
[8] Pardos A, et al. Gamification and Coaching in Remote Monitoring and Care Platforms. Studies in Health Technology and Informatics. 2022;294:644-8.
[9] https://www.heart.org/en/about-us
[10] https://www.sleepfoundation.org/
[11] Wooldridge M, Jennings NR. Intelligent agents: Theory and practice. The knowledge engineering review. 1995 Jun;10(2):115-52.
[12] Mutawa AM, Alzuwawi MA. Multilayered rule-based expert system for diagnosing uveitis. Artificial intelligence in medicine. 2019 Aug 1;99:101691.
[13] Hossain S, Sarma D, Chakma RJ, Alam W, Hoque MM, Sarker IH. A rule-based expert system to assess coronary artery disease under uncertainty. InInternational Conference on Computing Science, Communication and Security 2020 Mar 26 (pp. 143-159). Springer, Singapore.
[14] Christie GI, et al. Gamifying CBT to deliver emotional health treatment to young people on smartphones. Internet Interventions. 2019 Dec 1;18:100286.
[15] Cechetti NP, et al. Developing and implementing a gamification method to improve user engagement: A case study with an m-Health application for hypertension monitoring. Telematics and Informatics. 2019 Aug 1;41:126-38.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230130

MedSecurance Project: Advanced Security-for-Safety Assurance for Medical Device IoT (IoMT)

Parisis GALLOS[a,1], Rance DeLONG[b], Nicholas MATRAGKAS[c], Allan BLANCHARD[c], Chokri MRAIDHA[c] Gregory EPIPHANIOU[d], Carsten MAPLE[d], Konstantinos KATZIS[e], Jaime DELGADO[f], Silvia LLORENTE[f], Pedro MALÓ[g], Bruno ALMEIDA[g], Andreas MENYCHTAS[h], Christos PANAGOPOULOS[h], Ilias MAGLOGIANNIS[h], Petros PAPACHRISTOU[i], Mariana SOARES[j], Paula BREIA[j], Ana Cristina VIDAL[j], Martin RATZ[k], Ross WILLIAMSON[k], Eduard ERWEE[k], Lukasz STASIAK[k], Orfeu FLORES[l], Carla CLEMENTE[l], John MANTAS[a], Patrick WEBER[a], Theodoros N. ARVANITIS[m] and Scott HANSEN[b]

[a] *European Federation of Medical Informatics, Switzerland*
[b] *The Open Group, UK*
[c] *CEA, List, Université Paris-Saclay, France*
[d] *University of Warwick, UK*
[e] *European University Cyprus, Cyprus*
[f] *Universitat Politecnica de Catalunya, Spain*
[g] *Unparallel Innovation, Portugal*
[h] *BioAssist S.A., Greece*
[i] *HYGEIA Medical Group, Greece*
[j] *Centro Garcia de Orta, Hospital Garcia de Orta, Portugal*
[k] *Doccla AB, Sweden*
[l] *STAB VIDA, Portugal*
[m] *University of Birmingham, UK*

ORCiD ID: Parisis Gallos https://orcid.org/0000-0002-8630-7200, Rance DeLong https://orcid.org/0000-0002-6089-4734, Nicholas Matragkas https://orcid.org/0000-0002-8594-1912, Allan Blanchard https://orcid.org/0000-0001-7922-4880, Chokri Mraidha https://orcid.org/0000-0003-2993-5734, Dr Gregory Epiphaniou https://orcid.org/0000-0003-1054-6368, Konstantinos Katzis https://orcid.org/0000-0002-1470-2105, Jaime Delgado https://orcid.org/0000-0003-1366-663X, Silvia Llorente https://orcid.org/0000-0003-2000-6912, Pedro Maló https://orcid.org/0000-0001-6171-7345, Bruno Almeida https://orcid.org/0009-0001-0535-994X, Andreas Menychtas https://orcid.org/0000-0002-4510-5522, Christos Panagopoulos https://orcid.org/0000-0001-9282-0919, Ilias Maglogiannis https://orcid.org/0000-0003-2860-399X, Petros Papachristou https://orcid.org/0000-0001-5508-1934, Mariana Soares https://orcid.org/0000-0002-5538-334X, Paula Breia https://orcid.org/0000-0001-5188-9215, Ana Cristina Vidal https://orcid.org/0000-0002-3841-583X, Carla Clemente https://orcid.org/0000-0002-3384-4563, John

[1] Corresponding Author: Parisis Gallos, European Federation of Medical Informatics (EFMI), Ch de Maillefer 37, CH-1052 Le Mont-sur-Lausanne, Switzerland; E-mail: parisgallos@yahoo.com.

Mantas https://orcid.org/0000-0002-3051-1819, Patrick Weber https://orcid.org/0000-0003-4469-0464, Prof Theodoros N. Arvanitis https://orcid.org/0000-0001-5473-135X

Abstract. The MedSecurance project focus on identifying new challenges in cyber security with focus on hardware and software medical devices in the context of emerging healthcare architectures. In addition, the project will review best practice and identify gaps in the guidance, particularly the guidance stipulated by the medical device regulation and directives. Finally, the project will develop comprehensive methodology and tooling for the engineering of trustworthy networks of inter-operating medical devices, that shall have security-for-safety by design, with a strategy for device certification and certifiable dynamic network composition, ensuring that patient safety is safeguarded from malicious cyber actors and technology "accidents".

Keywords. Cyber Security, Medical Devices, IoT, Internet of Things

1. Introduction

According to the Annex I section 1 of the Medical Devices Regulation, both security and safety have to be considered for medical devices as patients' safety may be compromised due to "security issues" which may have "safety impacts" [1]. Weak security refers to security vulnerabilities that might be exploited to modify the normal behaviour of a medical device. On the other hand, restrictive security refers to very strict security measures that might affect the functional safety of a device [2]. Modern medical devices software development practices are the leading solutions to address complexity and evolution [3,4]. However, substantial challenges remain in achieving interoperability, dependability, and trustworthiness at scale within a diverse commercial medical device market facing an escalating threat environment. Meanwhile, advances in healthcare IT systems have resulted in complex socio-technical architectures, which deliver integrated and patient-centered services using medical devices. All these transformations, in addition to clinical benefits, they also introduce risks including security risks that need to be understood and managed to be reduced to acceptable levels [3]. There are numerous reports of new types of security vulnerabilities for this kind of architectures, which challenge the effectiveness of the current security tools [5-7].

MedSecurance envisages to address the identified challenges and go beyond the state-of-the-art by proposing a novel tool-supported methodology for safety-security co-analysis as part of a Threat Vulnerability Risk Analysis (TVRA) framework [8,9]. The proposed novel methodology will combine architectural and graph-based formalisms to model the system, possible attacks and failures, and how these propagate through the system components. The methodology will support the specification of the interdependencies between safety and security, and it will include interactions such as conflict between safety and security requirements, and conditional dependencies between the two. New metrics for quantifying the interaction of safety and security within a medical device will be proposed and they will support trade-off analysis.

2. MedSecurance Objectives and Approach

MedSecurance project will conceive novel methodologies, infrastructures, and technologies that enable an effective, harmonious and continuous development and evolution of secure system engineering management activities in Internet of Medical Things (IoMT). Project's main objective is to advance knowledge and basic understanding of decision making in diverse IoMT threat landscapes based on different system and component level interactions. This can be accomplished via the development of a novel holistic strategy that considers the interdependence of several IoMT subsystems, information exchange, risk thresholds, and regulatory ramifications. At this end, scalable and verifiable secure system engineering management solutions that capture, communicate, and act on these complexities in order to improve decision-making in cyber defense while automating cybersecurity assurance will be provided. Figure 1 illustrates the concept of MedSecurance.

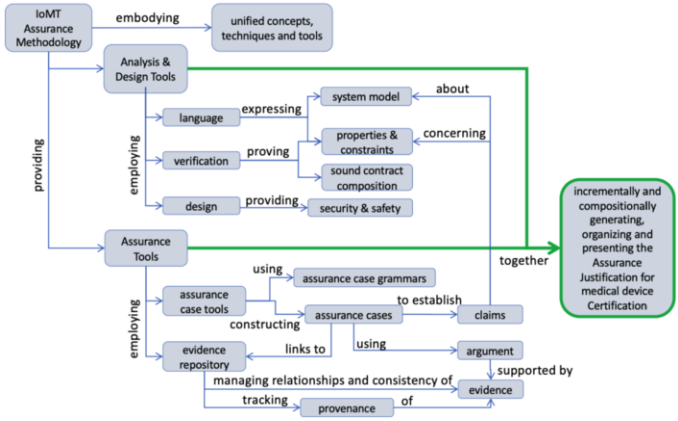


Figure 1. Concept of MedSecurance.

A stepwise approach will be used to achieve the project's main objective.

2.1. Systematic review, concept, and gap analysis of security approaches for the Internet of Medical Things (IoMT)

The project will perform a systematic review of security and safety standards and guidelines applicable to a) healthcare and b) health IT systems in general. The analysis will identify the main recommendations and concepts behind each standard and will perform a gap analysis with respect to the MDCG 2019-16 [10]. Furthermore, the standards and guidelines will be reviewed to identify gaps in the guidance with respect to the architectures and technologies identified in the first step. To proceed with Gaps mapping, a methodology that employs an iterative approach to reviewing and mapping the relevant documents and standards for inclusion in the full gap analysis process will be followed. Additionally, an elaboration and analysis of typical and alternative architectures for IoMT will be performed, to include system-driven risk-based threat modelling, vulnerability analysis, fault tree, Failure Mode and Effect Analysis (FMEA), and architectural patterns that substantially support specific classes of safety and security properties [11-13]. The purpose is to define and implement the appropriate TVRA processes and workflows based on best practices and compliant with existing legal and regulatory ramifications in both security engineering and resilience lines of effort.

2.2. Requirements and design of harmonized tools and methods for the unification of automated security and safety assurance for certification of IoMT

The project will examine modelling the integrated risk assessment approach proposed by ENISA [14], along with modelling the minimum viable security concepts required for assurance, which will be found in the literature. This will harmonize different security approaches and allow the transformation of terminology used in legacy certifications and the application of different standards. Assurance Automation design encompasses architectural, behavioural and communication modelling, semantic modelling (ontologies), modelling of essential characteristics, trust modelling of interfaces (contracts), as well as characterization of vulnerabilities analysis of design and implementation representations (design and code) to verify essential characteristics and marshalling of demonstrably sufficient evidence to support medical device safety/security certification (assurance cases).

2.3. Development of a security assurance automation toolbox

Develop assurance cases patterns and blueprints that are composable to demonstrate satisfaction of conformance with standards, regulations, legal obligations, and security-for-safety objectives by incorporating evidence from the architecture, design and implementation analyses of medical device connectivity solutions. In addition, the project will look at the interoperability software standards used in healthcare, and will implement interfaces that will assure the secure integration of components when their individual contracts are satisfied by their respective manufacturers. This will include developing of FHIR profiles for security assurance, and code security review and implementation of RESTful code (which is the main standard in healthcare). A related tool will allow the generation of secure code based on the different data exchange configurations.

2.4. Verification and Validation of the methods and tools by the Industry

Industry validation of new risk assessment and security assurance methods and tools will take place in the context of the MedSecurace. Pilot case evaluations by multiple Medical Devices suppliers under three project Use Cases for evaluating automated security assurance tools and methods. An appropriate architecture for implementing the process and enabling traceability between the system-level and component-level security requirements in IoMT via the programme will be developed.

2.5. MDCG 2019-16 recommendations, dissemination and engagement of stakeholders

The project will propose updates to the guidance that will bridge the gaps that will be identified. Furthermore, the project will expand the guidance, offering specific methods to be used (or references to standards) appropriate for each stage of the lifecycle and each architecture. Trade-off studies among alternative implementation technologies to inform choices will be needed to provide rationale for those choices. The proposed lifecycle and methods will correspond to a minimum assurance justification that will be identified by the prevailing certification authority. The project will incorporate a co-production approach identifying appropriate stakeholders who will offer knowledge and expertise,

including regulators, manufacturers as well as operators of medical devices and healthcare facilities.

3. Expected Outcomes - Conclusions

The MedSecurance provides a framework for ensuring proper security governance and empowering management to make security-aware choices about the evolving threats and risks in IoMT. A multi-layer Threat, Vulnerability and Risk Assessment (TVRA) approach will meant to be adaptable to any IoMT environment where the resilience lines of effort and security processes are entrenched but are specifically geared to risk identification, assessment, and treatment in order to allow the creation and management of security needs for such unique medical cyberinfrastructures. The proposed solutions will be co-developed and validated with our medical industry user partners, and complemented by engagement of healthcare industry stakeholders in support of the recommendations to existing guidelines that will also be developed in the project.

Acknowledgements

This work is co-funded by the HORIZON.2.1 - Health Programme of the European Commission, Grant Agreement number: 101095448 - Advanced Security-for-safety Assurance for Medical Device IoT (MEDSECURANCE).

References

[1] EU Regulation for Medical Devices. Available at: https://www.medical-device-regulation.eu/download-mdr/
[2] Katzis K, Jones RW, Despotou G. The Challenges of Balancing Safety and Security in Implantable Medical Devices. Stud Health Technol Inform. 2016;226:25-8.
[3] Despotou G, et al. A framework for synthesis of safety justification for digitally enabled healthcare services. Digit Health. 2017 Apr 24;3:2055207617704271.
[4] Czarnecki K. Requirements engineering in the age of societal-scale cyber-physical systems: The case of automated driving. In2018 IEEE 26th Intern Requirements Engineering Conf (RE) 2018 (pp. 3-4). IEEE.
[5] Yaacoub JP, Salman O, Noura HN, Kaaniche N, Chehab A, Malli M. Cyber-physical systems security: Limitations, issues and future trends. Microprocessors and microsystems. 2020 Sep 1;77:103201.
[6] Meng N, et al. Secure coding practices in java: Challenges and vulnerabilities. InProceedings of the 40th International Conference on Software Engineering 2018 May 27 (pp. 372-383).
[7] Ferrara P, Mandal AK, Cortesi A, Spoto F. Static analysis for discovering IoT vulnerabilities. International Journal on Software Tools for Technology Transfer. 2021 Feb;23:71-88.
[8] Threat, Vulnerability And Risk Assessment (TVRA). Available at: https://aipriskconsulting.com/theat-vulnerability-and-risk-assessment-tvra/
[9] Atay S, Masera M. Challenges for the security analysis of Next Generation Networks. Information security technical report. 2011 Feb 1;16(1):3-11.
[10] MDCG 2019-16 - Guidance on Cybersecurity for medical devices Available at: https://ec.europa.eu/docsroom/documents/41863.
[11] Ericson CA. Fault tree analysis. InSystem Safety Conference, Orlando, Florida 1999 (Vol. 1, pp. 1-9).
[12] Abdo H, et al. A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie–combining new version of attack tree with bowtie analysis. Computers & security. 2018 Jan 1;72:17595.
[13] Llorente S, Delgado J. Implementation of Privacy and Security for a Genomic Information System Based on Standards. J Pers Med. 2022 May 31;12(6):915. doi: 10.3390/jpm12060915.
[14] ENISA RM/RA Framework. Available at: https://www.enisa.europa.eu/topics/risk-management/current-risk/business-process-integration/the-enisa-rm-ra-framework?v2=1&tab=details

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230131

Leveraging Clinical Data Warehouses to Measure Impact of Update Prescription Guidelines of Polyvalent Immunoglobulins in 2018 in France

Morgane PIERRE-JEAN [a,1] , Guillaume BOUZILLE[a], Mathilde BORIES[b], Pascal LE CORRE[b,c,d] and Marc CUGGIA[a]
[a] *Univ Rennes, CHU Rennes, INSERM, LTSI-UMR 1099, Rennes, France*
[b] *Pôle Pharmacie, Service Hospitalo-Universitaire de Pharmacie, CHU de Rennes*
[c] *Laboratoire de Biopharmacie et Pharmacie Clinique, Faculté de Pharmacie, Rennes*
[d] *Univ Rennes, CHU Rennes, Inserm, EHESP, Irset UMRS 1085*

Abstract. In France and in other countries, we observed a significant growth in human polyvalent immunoglobulins (PvIg) usage. PvIg is manufactured from plasma collected from numeral donors, and its production is complex. Supply tensions have been observed for several years, and it is necessary to limit their consumption. Therefore, French Health Authority (FHA) provided guidelines in June 2018 to restrict their usage. This research aims to assess the guidelines' impact of the FHA on the use of PvIg. We analyzed data from Rennes University Hospital, where all PvIg prescriptions are reported electronically with quantity, rhythm, and indication. From the clinical data warehouses of RUH, we extracted comorbidities and lab results to evaluate the more complex guidelines. We globally noticed a reduction in the consumption of PvIg after the guidelines. Compliance with the recommended quantities and rhythms have also been observed. By combining two sources of data, we have been able to show an impact of FHA's guidelines on the consumption of PvIg.

Keywords. Retrospective study, Polyvalent Immunoglobulins, monitoring, clinical data warehouse

1. Introduction

Worldwide, there is a significant growth in PvIg usage. Therefore, supply tensions have emerged, for now, since several years [1]. In France because of blood product safety regulations, PvIg are submitted to a traceability procedure. The prescription and administration of the PvIg are therefore well controlled and monitored. However, this drug is frequently out of stock and subject to regular guideline usage to prioritize diseases and relevant clinical cases. In France, the FHA gave more restrictive new indications in June 2018. The second guideline update was performed in April 2019. In the context of monitoring the appropriate usage of PvIg consumption after these new guidelines, we conducted a retrospective study on PvIg in Rennes University Hospital (RUH). We assessed the impact of FHA guidelines at several levels. First, we measured consumption

[1] Corresponding Author: Morgane PIERRE-JEAN, E-mail: morgane.pierre-jean@chu-rennes.fr.

by patient-year, and then we evaluated the appropriateness of quantities and rhythms according to the guidelines.

2. Materials and Methods

At RUH, traceability procedure is conducted by pharmacists who fill a form for each PvIg prescription. For this study, we collected PVIg administrations for patients in RUH between 1 January 2013 and 31 December 2022. Patient identification, weight, PvIg dosage received, rhythms, and indication (pathology) were available through a database recording all filled PVIg prescription forms. We also used the clinical data warehouse of RUH to collect more data like comorbidities and lab results. We standardized the disease names and the received quantity of PvIg. Some guidelines require a threshold on a lab dosage or comorbidity such as kidney failure. We observed the evolution of the consumption of PvIg within the RUH at two levels; overall in the entire hospital and by ward (neurology, hematology, and internal medicine).

3. Results, Discussion and Conclusions

More than 1,500 unique patients received at least one dose of PvIg at RUH between 2013 and 2020. The number of prescribed doses per year declined after the publication of the guidelines in June 2018 from 2,465 to 2,276. For Hematology and internal medicine, we noticed a decrease in both the treated patients and the quantity of PvIg just before the first guidelines (probably because of out-of-stock). For neurology, the number of patients who have received PvIg is growing, but the quantity is stable. For all wards, the consumption is growing up since 2022. For a secondary criterion, we achieved cross databases (PVIg prescription database of pharmacists and clinical data warehouse of RHU). We found a better compliance to guidelines on this criterion after guidelines (78% of compliance before guidelines to 86% after guideline). The impact of guidelines on diseases with a small number of patients is complex to assess. Overall, the impact of guidelines on the drop in PvIg consumption is difficult to assess because of several other factors, such as out-of-stock periods and the COVID-19 pandemic. Similar study was recently performed in Catalonia (Spain) [2]. They observed that the mean consumption decrease between 2020 and 2021 but it is probably due to COVID-19 (not observed in our study). It could be interesting to expand the study to a multi-centric analysis by using clinical data warehouse in other hospitals to bring a better overview of pathologies with a small number of patients.

Acknowledgements: This work is supported by Epi-phare (Groupement d'intérêt scientifique).

References

[1] Yailian AL, et al. Parcours de soins du patient et sécurisation des approvisionnements en immunoglobulines dans les déficits immunitaires primitifs: résultats d'une étude pilote. InAnnales Pharmaceutiques Françaises 2021 May 1 (Vol. 79, No. 3, pp. 301-307). Elsevier Masson.

[2] Solís-Díez G, et al. Dealing With Immunoglobulin Shortages: A Rationalization Plan From Evidence-Based and Data Collection. Front Public Health. 2022 May 19;10:893770.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230132

Electronic Health Records as Information Source in Assessment of the Effectiveness of Delivered Care - A Pilot Study

Hanna VON GERICH[a,b,1], Hans MOEN[c] and Laura-Maria PELTONEN[b]
[a] *Turku University Hospital, Turku, Finland*
[b] *Department of Nursing Science, University of Turku, Turku, Finland*
[c] *Department of Computer Science, Aalto University, Espoo, Finland*
ORCiD ID: Hanna von Gerich https://orcid.org/0000-0003-1036-2163
Hans Moen https://orcid.org/0000-0003-1418-7892
Laura-Maria Peltonen https://orcid.org/0000-0001-5740-6480

Abstract. Effectiveness is a key element of high quality health services. The aim of this pilot study was to explore the potential of electronic health records (EHR) as an information source for assessing the effectiveness of nursing care by investigating the appearance of nursing processes in the documentation of care. Deductive and inductive content analysis were used in a manual annotation of ten patients' EHRs. The analysis resulted in the identification of 229 documented nursing processes. The results indicate that EHRs can be used in decision support systems for assessing effectiveness of nursing care, however, future work is needed to verify these findings in a larger data set and extend to other dimensions related to care quality.

Keywords. Electronic health records, hospital, nursing, process assessment

1. Introduction

Effectiveness is one key element in providing high quality health services, defined as delivering evidence-based care to those in need [1]. The aim of this pilot study was to explore the potential of electronic health records (EHR) as an information source for assessing the effectiveness of nursing care by investigating the appearance of evidence-based nursing processes in the documentation of care. The results can be utilized as an element of nursing care quality evaluation when developing decision support systems (DSS) for intelligent evaluation of nursing care for nursing leadership and healthcare management.

[1] Corresponding Author: Hanna von Gerich, Department of Nursing Science 20014 University of Turku, Turku, Finland.; E-mail: hanna.m.vongerich@utu.fi

2. Methods

This study was a retrospective descriptive pilot study using EHR data collected in a cardiac center of one hospital district in Finland. The data set included structured and free text nursing notes. The patient record sample (n=10) was selected using random sampling. In total, 585 pages containing 198793 words were analyzed. A manual annotation of the EHRs was guided by the nursing process condensed into three steps: 1) assessing the patient's need, 2) provision of nursing intervention and 3) assessment of patient outcome. The collected nursing processes were organized using deductive and inductive content analysis methods. The study followed the European Code of Conduct for Research Integrity guidelines and ethical review was done.

3. Results

The total amount of expressions annotated was 3636, out of which 1914 (52.7%) were derived from free text entries. In total, 229 nursing processes were identified, out of which 144 (62.9%) included all three steps of the nursing process. The inductive analysis identified nursing processes that can be divided into two main categories: 1) *Responding to the patient's long term needs,* and 2) *Responding to the patient's imminent needs.* The majority (n=213, 93.0%) of the processes in both categories were related to responding to the physical care needs of the patient.

4. Discussion

The identified nursing processes mainly described nurses' responding to physical care needs, with very scarce descriptions of the psychosocial and relational elements of care. Over one third of the nursing processes lacked a clear description of the evaluation of the patient outcomes. This pilot study shows promising results when it comes to identifying nursing care effectiveness from analyzing nursing documentation in EHRs. This also means that this task has the potential to be automatized in the form of a machine learning-based DSS. However, as a first step, careful consideration of how to systematically evaluate the documented content and its quality is needed. Limitations of this study include a small sample and using data from only one cardiac center.

5. Conclusions

EHRs show potential as an information source in the development of DSS for nursing and healthcare management in the assessment of the effectiveness of nursing care. However, consideration should be given on developing methods for assessment of documentation quality, as this will impact the output given by the system.

References

[1] World Health Organization, Organization for Economic Co-operation and Development, World Bank. Delivering quality health services: a global imperative for universal health coverage. WHO; 2018.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230133

What Are We Talking About When We Talk About Information-Driven Care? A Delphi-Study on a Definition

Elin SIIRA[a,b,1], Petra SVEDBERG[a], Carl SAVAGE[a,c] and Jens NYGREN[a]
[a]*School of Health and Welfare, Halmstad University, Sweden*
[b]*Centre for Person-Centred Care (GPCC), University of Gothenburg, Sweden*
[c] *Medical Management Centre, Karolinska Institutet, Sweden*
ORCiD ID: Elin Siira https://orcid.org/0000-0002-3097-9147

Abstract. In Sweden, the term *information-driven care* has recently been put forward by healthcare organizations and researchers as a means for taking a comprehensive approach to the introduction of Artificial Intelligence (AI) in healthcare. The aim of this study is to systematically generate a consensus definition of the term information-driven care. To this end, we are conducting a Delphi study utilizing literature and experts' opinions. The definition is needed to enable knowledge exchange on information-driven care and operationalize its introduction into healthcare practice.

Keywords. Information-driven care, information driven care, artificial intelligence, Delphi study

1. Introduction

New ways of providing and organizing care by the means of deploying Artificial Intelligence (AI) hold great potential to alleviate contemporary healthcare challenges [1]. While the technical aspects of AI are often highlighted, the realization of its potential in healthcare depends on its implementation and integration into everyday practice [2]. For a successful utilization of AI in healthcare, there is a need to advance theory and empirical evidence regarding its deployment by bringing together insights from research and practice focusing on both technical, health, and social aspects [3].

In Sweden, the use of the term *information-driven care* has recently been put forward by Halmstad University and Region Halland as a means for taking a comprehensive approach to the utilization of AI in healthcare [4]. The idea behind the term holds potential, but to date, there are no research publications stating what 'information-driven care' is. To operationalize the term – to go from hype and hope to usefulness in practice – a clarification of its meaning is needed. This study aims to develop a systematically generated consensus on a definition of the term based on current literature and the views of leaders and strategists in Swedish healthcare.

[1] Corresponding author: Elin Siira, Halmstad University, Box 823, 301 18 Halmstad, Sweden, E-mail: elin.siira@hh.se.

2. Methods

To systematically generate a consensus on the definition of the term information-driven care based on the views of leaders and strategists in Swedish healthcare we will conduct a Delphi-study [5]. The participants will be recruited through a national network on developing information-driven care in practice. The study will be reported according to the Conducting and Reporting of Delphi Studies (CREDES) standard [6]. For the purpose of the Delphi-study we will carry out a preparatory literature review.

3. Results

This study is currently ongoing. The results will provide clarification on the definitions of information-driven care. Furthermore, the results will reflect experts' opinion on the definition of information-driven care.

4. Discussion

Successful operationalization and knowledge exchange concerning terms or concepts in research demand that they are defined and described. For the use of the term information-driven care to support introduction of AI in healthcare it needs to be more well defined. The current work addresses this issue.

5. Conclusion

This paper presents the ongoing work of a Delphi-study for generating a consensus on a definition of the term information-driven care. To our knowledge, it is the first study that addresses the definition of the term information-driven care.

References

[1] Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. JBI. 2019 Dec;100:103311: 1-14, doi: https://doi.org/10.1016/j.jbi.2019.103311

[2] Nilsen P, Reed J, Nair M, Savage C, Macrae C, Barlow J, Svedberg P, Larsson I, Lundberg L, Nygren J. Realizing the potential of artificial intelligence in healthcare: Learning from intervention, innovation, implementation and improvement sciences. Front. Health Serv. 2022 Sept; 2:1-8, doi: 10.3389/frhs.2022.961475

[3] Svedberg P, Reed J, Nilsen P, Barlow J, Macrae C, Nygren J. Toward Successful Implementation of Artificial Intelligence in Health Care Practice: Protocol for a Research Program. JMIR Res. Protoc. 2022 Mar;11(3):e34920: 1-13, doi: doi:10.2196/34920

[4] Halmstad University. IDC – Information Driven Care 2022 [Retrieved December 2022 from: https://www.hh.se/english/research/research-programmes/idc---information-driven-care.html.]

[5] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000 Oct;32(4):1008-15, doi: 10.1046/j.1365-2648.2000.t01-1-01567.x

[6] Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. Palliat Med. 2017 Sep;31(8):684-706. doi: 10.1177/0269216317690685

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230134

Identifying and Predicting Postoperative Infections Based on Readily Available Electronic Health Record Data

Siri Lise van der MEIJDEN[a,b,1,2], Anna van BOEKEL[b], Laurens SCHINKELSHOEK[a,2], Harry van GOOR[c], Mark de BOER[b], Ewout STEYERBERG[b], Bart GEERTS[a,2] and Sesmu ARBOUS[b]

[a] *Healthplus.ai, Amsterdam, The Netherlands*
[b] *Leiden University Medical Center, Leiden, The Netherlands*
[c] *Radboud University Medical Center, Nijmegen, The Netherlands*

ORCiD ID: S.L. van der Meijden https://orcid.org/0000-0002-9343-0899, A.M. van Boekel https://orcid.org/0000-0003-0774-0995, L.J. Schinkelshoek https://orcid.org/0000-0002-3112-5231, H. van Goor https://orcid.org/0000-0003-0323-4876, E.W. Steyerberg https://orcid.org/0000-0002-7787-0122, B.F. Geerts https://orcid.org/0000-0002-0210-7202, M.S. Arbous https://orcid.org/0000-0001-5242-3257

Abstract. Identification of postoperative infections based on retrospective patient data is currently done using manual chart review. We used a validated, automated labelling method based on registrations and treatments to develop a high-quality prediction model (AUC 0.81) for postoperative infections.

Keywords. Artificial Intelligence, Prediction, Postoperative infections, Electronic Health Record

1. Introduction

Postoperative infections are common complications with a global reported incidence of 9.0% [1]. To allow the development of prediction models for clinical decision support, we need to identify which patients had a postoperative infection in retrospective data. This process is called labelling, and is often done by manual review of the Electronic Health Record (EHR) as complications are severely under-registered [2]. Automation of this labour-intensive process is needed to enhance surveillance and scalability of prediction models. We aimed to develop and validate a postoperative infection risk prediction model using a domain knowledge-based labelling method based on readily available, non-free text, EHR patient data.

[1] Corresponding author: Siri Lise van der MEIJDEN, E-mail: S.L.van_der_meijden@lumc.nl
[2] S.L. and L.J. are employees and B.F. is owner and shareholder of Healthplus.ai

2. Methods

Adult surgical patients admitted to the Leiden University Medical Center (LUMC) between 2012 and 2022 were included for model development. In consultation with clinicians and data scientists, a definition was determined to label all postoperative bacterial infections that required complication or diagnosis registration, pharmacological treatment (initiating at least 24h after surgery for a minimum duration of 72 hours, excluding prophylactic regimes) and/or surgical intervention to treat infections. The dataset was split into a development part (2012-2020) and a temporal test part (2021-2022). Thirty cases with an infection and 30 without an infection according to the labelling method were randomly selected from the test dataset. Two clinicians independently determined for each case whether a postoperative infection occurred based on medication prescriptions, procedure information and complication and diagnosis registrations extracted from the EHR.

After this label validation step, an XGBoost model was trained on the development dataset to predict any bacterial infection within 30 days of surgery according to the definition. Medication prescriptions, patient characteristics, vital functions, comorbidities, and procedure characteristics were used as input features.

3. Results, Discussion and Conclusions

This research was done under the General Data Protection Regulation and a waiver for medical ethical approval from the LUMC was obtained (G18.129). Average overall infection rates were 12% (n=7,093 out of 59,106 procedures) in the development dataset and 13% (n=1,264 out of 9,722 procedures) in the test dataset.The automated infection label had 93% (54/60) agreement compared to manual labelling. We were able to predict postoperative infections on the test dataset with an area under the receiver operating characteristic curve of 0.81 (95% confidence interval (CI) 0.80-0.83), calibration slope of 0.81 (95% CI 0.77-0.86), and a positive net benefit, a measure that indicates clinical usefulness, for a broad range of threshold probabilities (0-85%).

The development of high-quality prediction models for clinical decision support requires reliable automatization of data processing and labelling. A treatment-based approach is needed to identify infections, since a definition based on solely registered complications would under estimate the risk. The next steps are to further validate the infection labelling method, improve the prediction model by adding more features to the dataset and externally validate the model on different hospital datasets to allow safe and broader implementation across hospital systems.

It is feasible to predict postoperative infections within 30 days of surgery with acceptable performance using an automated infection labelling method based on registrations and treatments of infection.

References

[1] International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. Br J Anaesth. 2016 Oct 31;117(5):601-609.
[2] Ubbink DT, Visser A, Gouma DJ, Goslings JC. Registration of surgical adverse outcomes: a reliability study in a university hospital. BMJ Open. 2012 May 25;2(3):e000891.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230135

Improving Pressure Ulcers Prediction in Nursing Homes with ML Algorithm

Clara CHARON [a,b,1], Pierre-Henri WUILLEMIN [b] and Joël BELMIN [c,d]
[a] *Teranga Software, Paris, France*
[b] *LIP6 (UMR 7606), Sorbonne Université, Paris, France*
[c] *AP-HP, Hôpital Charles-Foix, Ivry-sur-Seine, France*
[d] *LIMICS (UMR 1142), Sorbonne Université, Paris, France*

Abstract. An automated ML classifier predicting pressure ulcers one-month before performs better than the reference methods currently used in nursing homes.

Keywords. Machine Learning, Nursing Homes, Braden Scale, Pressure Ulcers

1. Introduction and Methods

As the population ages, the prevalence of chronic disease and disability increases, and the number of older adults living in nursing homes (NHs) grows. NH residents are a high-risk population for developing pressure ulcers (PU), which are skin injury related to immobility and ischemia induced by prolonged pressure and microvascular compression. This condition causes pain, infection and impaired quality of life. Treatment of PU is difficult, expensive and time consuming for nurses, but it is highly preventable with a specific, multidisciplinary approach [1]. A large database of electronic health records (EHR) from a NH software[2] allowed us to create a classifier made using machine learning (ML) and specifically Bayesian networks (BN) [2], that accurately predicts PU based on data available one month prior to PU onset [3]. To assess the risk of PU, NH staff currently use clinical tools such as the Braden or Norton scales which are considered the reference, even if they have been criticized [4]. We identified residents in the database who had a risk assessment for PU by the Braden scale (which was more widely used than the Norton scale). It gave us the opportunity to compare its finding with our BN classifier calculated at the same time and sample and in the rest of the residents. The risk prediction by the Braden scale (with a cut-off of 12) and the BN classifier were compared with the presence/absence of PUs one month later.

2. Results and Discussion

Among the 37,231 residents analysed in our database, only 2,110 (5.6%) had a Braden scale PU risk assessment. Among them, 695 residents (33%) suffered from PU a month

[1] Corresponding Author : Clara Charon, LIP6, 4 Place Jussieu, 75005 Paris; E-mail: clara.charon@lip6.fr.
[2] This research is funded by Teranga Software and the ANRT CIFRE Grant n°2019/1519 for CC and was approved by the CESREES of the Health Data Hub (#9458633).

later. This contrasted with the lower PU occurrence (5,656; 16%) among the 35,121 residents which were not assessed by the Braden scale, indicating that the use of the Braden scale was yet an indicator of a high-risk for PU, independently of its result. It is likely that NH staff realized PU risk assessment by the Braden scale on the basis of their clinical awareness. Sensitivity, specificity, positive and negative predictive values based on the confusion matrices of risk assessments by the Braden scale and the BN classifier obtained in the residents assessed by the Braden scale (higher-risk sample) as well as the BN classifier in the rest of the NH residents (lower-risk sample) are shown in Table 1.

Table 1. Scores of Braden Scale and BN Classifier in the higher-risk (HR) and in the lower-risk (LR) sample.

	Braden scale	BN Classifier (HR)	BN Classifier (LR)
Sensitivity	0.15	0.99	0.86
Specificity	0.91	0.01	0.91
Positive predictive value	0.46	0.33	0.65
Negative predictive value	0.69	0.83	0.97

The performance of the Braden scale was poor, as its sensitivity was 0.15, indicating that the scale strongly underestimated the risk of PU. The performance of the BN classifier in the lower-risk sample of residents was clearly better, with higher positive and negative predictive values. However, in the higher-risk sample, the BN classifier had low specificity and positive predictive value, indicating that the clinical awareness of the nurses who decided to perform the Braden scale captured most of the information provided by the BN classifier. These results highlight the importance of the clinical awareness of healthcare professionals, but it should be noted that their determinants are not known and cannot be easily translated into decision rules. The main limitation of our study is related to the rather small percentage of residents assessed by the Braden scale, selected on the basis of the clinical awareness of the nursing staff. In addition, we cannot exclude that in residents with a positive Braden scale score, nurses implemented care processes to prevent PU and, therefore, reduce its incidence. Thus, a better comparison of the performance of the Braden scale and the ML algorithms would have been better achieved by assessing the risk of PU by both methods on a random sample.

3. Conclusion

For PU risk assessment, the BN classifier performed much better than the gold standard: the Braden scale, which had low sensitivity and underestimated PU risk in a large proportion of residents. We observed that the nurses' clinical awareness that led them to perform a risk assessment by the Braden scale has an interesting predictive value which is found almost perfectly by the BN classifier. Our method had also fine results on the lower-risk sample, making it a legitimate support in the decision of the population at risk.

References

[1] Belmin J, et al. Gériatrie: pour le praticien. Elsevier Health Sciences; 2016 Sep 7.
[2] Ducamp G, et al. aGrUM/pyAgrum: a toolbox to build models and algorithms for Probabilistic Graphical Models in Python. InInternational Conference on Probabilistic Graphical Models 2020 Feb 2. PMLR.
[3] Charon C, et al. Learning Bayesian Networks for the Prediction of Unfavorable Health Events in Nursing Homes. InChallenges of Trustable AI and Added-Value on Health 2022 (pp. 147-148). IOS Press.
[4] Chen HL, Shen WQ, Liu P. A meta-analysis to evaluate the predictive validity of the braden scale for pressure ulcer risk assessment in long-term care. Ostomy Wound Manage. 2016 Sep 1;62(9):20-8..

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230136

Domain Knowledge-Driven Generation of Synthetic Healthcare Data

Atiye Sadat HASHEMI [a], Amira SOLIMAN [a], Jens LUNDSTRÖM [a] and Kobra ETMINANI [a,1]
[a] *Center for Applied Intelligent Systems Research in Health, Halmstad University, Sweden*

Abstract. Healthcare longitudinal data collected around patients' life cycles, today offer a multitude of opportunities for healthcare transformation utilizing artificial intelligence algorithms. However, access to "real" healthcare data is a big challenge due to ethical and legal reasons. There is also a need to deal with challenges around electronic health records (EHRs) including biased, heterogeneity, imbalanced data, and small sample sizes. In this study, we introduce a domain knowledge-driven framework for generating synthetic EHRs, as an alternative to methods only using EHR data or expert knowledge. By leveraging external medical knowledge sources in the training algorithm, the suggested framework is designed to maintain data utility, fidelity, and clinical validity while preserving patient privacy.

Keywords. Domain Knowledge, EHR, Synthetic Data, Representation Learning

1. Introduction and Methods

Generating synthetic data, that is not collected from real-world occurrences but is artificially generated, is considered nowadays as an alternative to making data sharable whilst maintaining the constraints of data privacy and sensitivity [1]. In the healthcare domain, in particular, the availability of high-quality synthetic data can open up opportunities to improve healthcare quality and efficiency, policy evaluation, and large-scale biomedical research investigations. State-of-the-art techniques provide methodologies to generate synthetic electronic health records (EHRs) [2]. Yet, several challenges remain due to the longitudinal and temporal aspects, data heterogeneity, sparsity, skewed distributions, ensuring privacy as well as considering clinical knowledge in the process of modeling healthcare data. The focus of this study is to investigate the integration of domain knowledge to support greater patient-centered outcomes that are close to real clinical data by preserving relationships, distributions, predictive capabilities, and patients' privacy.

Figure 1 illustrates the proposed framework which combines the suggested building blocks including representation learning, generative adversarial network (GAN), post-hoc clinical evaluations, and external domain knowledge sources. Medical ontologies (e.g., International Classification of Diseases (ICD), Anatomical Therapeutic Chemical Classification System (ATC), clinical guidelines, Unified Medical Language System (UMLS), and SNOMED CT) provide powerful sources of information for understanding

[1] Corresponding Author: Kobra Etminani, E-mail: Kobra.Etminani@hh.se.

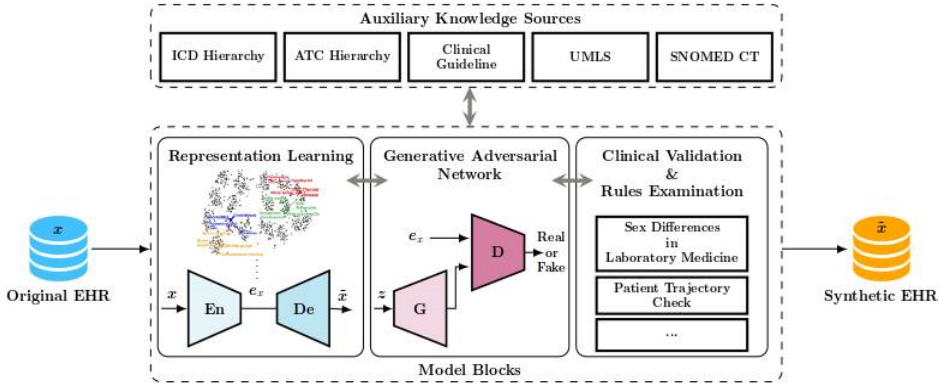


Figure 1. Our proposed data synthesizer framework.

disease progression and enriched categorization of diseases as well as medical treatments. These knowledge sources are integrated for enhancing the conditional representation space for the synthesizer, also allowing us to combine statistical models with data-driven models. The framework uses an Autoencoder as the representation learning module, then a GAN is used for generating data while we have the clinical validity as a ruling-out mechanism. We aim to integrate state-of-the-art representation learning techniques such as graph neural networks (GNNs) and graph convolutional transformers (GCTs) to extract latent structures and relations from EHR data [3].

2. Results, Discussion and Conclusions

While generating realistic data is challenging for researchers, domain knowledge-driven training schemes would be a promising solution. In this paper, we propose a synthesizer that employs a domain knowledge source combined with a deep learning-based model for generating longitudinal EHRs. Although there is some EHR generative software, we aim at considering the clinical knowledge in the synthesizer training process to address the gaps in the multidisciplinary aspects of medical and data science. We hope to outperform previous methods in terms of fidelity and privacy aspects. For this aim, not only do we focus on patients' training data, but also, we leverage diseases-based general knowledge to define several conditions in the training phase (based on the data of respective patient cohorts e.g., heart failure, chronic kidney disease, etc.,) to generate realistic privacy preserved EHRs. In the end, the synthesized EHR will be evaluated via relevant criteria such as fidelity, utility, privacy, and clinical validation metrics.

References

[1] Hernandez M, Epelde G, Alberdi A, Cilla R, Rankin D. Synthetic data generation for tabular health records: A systematic review. Neurocomputing. 2022 Apr 13.

[2] Goncalves A, Ray P, Soper B, Stevens J, Coyle L, Sales AP. Generation and evaluation of synthetic patient data. BMC medical research methodology. 2020 Dec;20(1):1-40.

[3] Chen J, Guo C, Lu M, Ding S. Unifying Diagnosis Identification and Prediction Method Embedding the Disease Ontology Structure From Electronic Medical Records. Frontiers in Public Health. 2022 Jan 20;9:793801.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230137

Learning from Health Professionals: A User-Centred Approach to Design a Wound Monitoring Platform

Beatriz FÉLIX[a,1], Ricardo MELO[a], Raquel MARQUES[b], Paulo ALVES[b] and Pedro Faria SALGADO[c]
[a] *Fraunhofer Portugal AICOS*
[b] *Centre for Interdisciplinary Research in Health, Institute of Health Sciences, UCP*
[c] *Health Solutions Manager, F3M Information Systems, SA*

Abstract. Health professionals are able to improve the care quality of chronic wounds by monitoring and reporting the wound status. Resorting to visual representations of wound status enhances comprehension by facilitating knowledge transfer to all stakeholders. However, selecting appropriate healthcare data visualisations is a critical challenge and healthcare platforms must be designed to meet their users' needs and constraints. This article describes the methods used to identify the design requirements and inform the development of a wound monitoring platform through a user-centred approach.

Keywords. Wounds and Injuries, User-Centred Design, Data Visualisation, eHealth

1. Introduction

Chronic wounds are a global problem that impact both healthcare and people's quality of life [1]. Characterization and reporting of wound condition are required to make the right diagnosis, gather and compare data, monitor healing progress, and determine therapy effectiveness [2]. To optimise comprehension, data should be presented as visual information [3]. But choosing visualisations for health professionals can be a challenge since they need simple and efficient data organisation to enable pattern detection [4].

As such, we followed a user-centred approach to learn about best practices and key metrics for wound treatment and monitoring, as well as to understand the users, their roles, and their needs. The end goal of this ongoing work is to create a management dashboard platform for health professionals, that aggregates and displays data retrieved from a wound-monitoring mobile application. Through applying these user-centred methods we are able to design a platform that fits the users' real-world requirements.

2. Methods

Our approach began with conducting semi-structured interviews. We inquired participants about their work methods and practices. This was followed by two assessment sessions of a dashboard from an available product that served as the basis for

[1] Corresponding Author: Beatriz Félix, E-mail: beatriz.felix@fraunhofer.pt

our work. Following these activities, we prototyped visualisations and integrated them into a new dashboard. Subsequently, we conducted guerrilla tests, where participants were given a tour of the prototype and requested to provide qualitative feedback. Finally, we conducted 5 usability tests, including a System Usability Survey where participants were given three tasks to complete in the prototype. All activities were conducted remotely, and all participants provided written informed consent.

Table 1. Methods, participants, and goals of each user research activity.

Method	Participants	Goal
Interviews	2 nurses, 2 clinical managers, 1 informal caregiver	Identify system actors, guidelines, and best practices in wound monitoring.
Evaluation sessions	2 head-nurses, 1 clinical manager	Understand the application's issues and assess the clarity and usability of each visualization.
Guerilla tests	2 homecare nurses, 2 clinical nurses, 1 head-nurse, 1 clinical manager	Identify main usability issues in the prototype.
Usability tests + SUS	1 homecare nurse, 1 clinical nurse, 1 head-nurse, 2 clinical managers	Test the prototype and assess self-perceived usability and satisfaction.

3. Preliminary Findings and Future Work

As preliminary findings of this exploratory phase of the work, we identified that health professionals require a platform that allows for personalised profiles as well as adding or modifying said profiles. We also found that information should be changed to account for diverse clinical contexts, as failing to do so can lead to low system adoption. Additionally, information should be current, relevant, and applicable to real-world practice. As for limitations, we could not observe in-person users' interaction with the platform. This will be addressed in future work, with the development of the platform and its assessment with representative users.

Acknowledgements

This work was done under the scope of "Wound Analysis to Support Clinical Decision" project (POCI-01-0247-FEDER-048922), Portugal 2020, co-funded by the European Regional Development Fund from European Union, framed in the COMPETE 2020.

References

[1] Sen CK, Gordillo GM, Roy S, Kirsner R, Lambert L, Hunt TK, et al. Human skin wounds: A major and snowballing threat to public health and the economy. Wound Repair and Regeneration 2009;17:763–71. https://doi.org/10.1111/j.1524-475X.2009.00543.x.

[2] Smet S, Probst S, Holloway S, Fourie A, Beele H, Beeckman D. The measurement properties of assessment tools for chronic wounds: A systematic review. International Journal of Nursing Studies 2021;121:103998. https://doi.org/10.1016/j.ijnurstu.2021.103998.

[3] Jayasinghe S. Describing complex clinical scenarios at the bed-side: Is a systems science approach useful? Exploring a novel diagrammatic approach to facilitate clinical reasoning. BMC Med Educ 2016;16:264. https://doi.org/10.1186/s12909-016-0787-x.

[4] Shneiderman B, Plaisant C, Hesse BW. Improving Healthcare with Interactive Visualization. Computer 2013;46:58–66. https://doi.org/10.1109/MC.2013.38.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230138

Patients' Experiences of Unwanted Access to Their Online Health Records

Annika BÄRKÅS[a,1], Anna KHARKO[a], Rose-Mharie ÅHLFELDT[b] and Maria HÄGGLUND[a]
[a] *Department of Women's and Children's Health, Uppsala University, Sweden*
[b] *School of Informatics, University of Skövde, Sweden*
ORCiD ID: Annika Bärkås https://orcid.org/0000-0002-1209-7714
Anna Kharko https://orcid.org/0000-0003-0908-6173
Rose-Mharie Åhlfeldt https://orcid.org/0000-0002-8607-948X
Maria Hägglund https://orcid.org/0000-0002-6839-3651

Abstract. Patient-Accessible Electronic Health Records (PAEHR) are particularly disputed in mental healthcare. We aim to explore if there is any association between patients having a mental health condition and someone unwanted seeing their PAEHR. A chi-square test showed a statistically significant association between group belonging and experiences of someone unwanted seeing their PAEHR.

Keywords. mental health, patient-accessible electronic health record (PAEHR)

1. Introduction

Patient-Accessible Electronic Health Records (PAEHRs) are secure web-based portals that are becoming more widespread but continue facing resistance in mental healthcare. Healthcare professionals (HCP) are concerned that mental health patients may be harmed by the content of their PAEHRs, while patients, in contrast, report benefits from reading their psychiatric records. Mental health is often considered particularly sensitive and stigmatized, and patients are often considered vulnerable and unable to manage their own care [1]. Patients with mental health conditions may therefore be more concerned about confidentiality and privacy issues related to their PAEHRs. This study aims to explore if there is an association between patients with mental health conditions and the experiences of someone unwanted seeing their PAEHR.

2. Methods

This 46-item survey was designed within the research project NORDeHEALTH (NordForsk Project #100477). The survey was distributed through the Swedish PAEHR platform 1177 for three weeks in January 2022. The survey responses were anonymous. This paper focuses on two participant groups; the Mental Health Group (MHG), which

[1] Corresponding Author: Annika Bärkås, Dag Hammarskjölds väg 14b, 1 floor, 751 85 Uppsala, Sweden. E-mail: annika.barkas@kbh.uu.se

includes respondents self-reporting experiences with mental healthcare, and the General Group, in which participants reported any other healthcare. Data analysis was performed in Jasp (v0.16.2) and included descriptive statistics and a chi-square test.

3. Results

The total number of respondents included in this analysis was 12,334. A chi-square test showed a statistically significant association between group belonging and the experiences of someone seeing their PAEHR without consent ($\chi 2(2, N=12334) = 234.57$, $p < .001$). In both groups, the majority responded that they have not experienced someone unwanted seeing their PAEHR; however, the proportion of those who had experienced this was larger in the MHG (see Table 1).

Table 1. Responses to the item "Have you experienced that someone has seen your health record that you did not want to share?"

Survey Item	MHG (*n*=3131)	General Group (*n*=9203)	*P* value
Has anyone seen your PAEHR?			< .001
Yes	256 (8.18%)	245 (2.66%)	
No	1894 (60.49%)	6539 (71.05%)	
Don't know/don't remember	981 (31.33%)	2419 (26.28%)	

4. Discussion

Respondents in both groups stated they had experienced someone seeing their PAEHR against their wish; however, the proportion was higher in the MHG. One could hypothesize that this is either due to more sensitive content in the MHG's records (hence less comfortable sharing), or that this group is viewed as more vulnerable and unable to care for themselves so others are more likely to try to access their records without their consent. Previous research indicates that the Swedish population trusts the healthcare system's ability to protect their information in the PAEHRs [2]; further research is needed to explain why patients in mental healthcare experienced more unwanted access to their health records.

References

[1] Schwarz J, Bärkås A, et al. Sharing Clinical Notes and Electronic Health Records with People Affected by Mental Health Conditions: Scoping Review. JMIR Ment Health. 2021 Dec;8(12):e34170.

[2] Belfrage S, Helgesson G, et al. Trust and digital privacy in healthcare: a cross-sectional descriptive study of trust and attitudes towards uses of electronic health data among the general public in Sweden. BMC Med Ethics. 2022 Mar;23(1):19.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230139

RD-MON - Building a Rare Disease Monitor to Enhance Awareness for Patients with Rare Diseases in Intensive Care

Romina BLASINI[a], Achim MICHEL-BACKOFEN[b], Henning SCHNEIDER[a] and Kurt MARQUARDT[b]
[a] *Department of medical informatics, University of Gießen, Germany*
[b] *Department of clinical and administrative data processing, University Hospital of Gießen and Marburg, site Gießen, Germany*

Abstract. Rare diseases are commonly defined by an incidence of less than 5/10000 inhabitants. There are some 8000 different rare diseases known. So even if a single rare disease is seldom, together they pose a relevant problem for diagnosis and treatment. This is especially true if a patient is treated for another common disease. University hospital of Gießen is part of the CORD-MI Project on rare diseases within the German Medical Informatics Initiative (MII) and a member of the MIRACUM consortium within the MII. As part of the ongoing Development for a clinical research study monitor within the use case 1 of MIRACUM, the study monitor has been configured to detect patients with rare diseases during their routine clinical encounters. The goal was to send a documentation request to the corresponding patient chart within the patient data management system for extended disease documentation to enhance clinical awareness for the patients' potential problems.

The project was started in late 2022 and has so far been successfully tuned to detect patients with Mucoviscidosis and place notifications within the patient chart of the patient data management system (PDMS) on intensive care units.

Keywords. Rare diseases, mucoviscidosis, study monitor, patient data management system, FHIR, OMOP

1. Introduction

Rare diseases are commonly defined by an incidence of less than 5 per 10000 inhabitants [1]. Rare diseases may pose problems if they go unnoticed during the treatment of other patient problems e.g., during a necessary surgical treatment with postoperative intensive care therapy or as part of a COVID 19 related intensive care treatment.
The University Hospital of Gießen (UHG) is part of the CORD-MI project on rare diseases in the German Medical Informatics Initiative (MII) [2] and a member of the MIRACUM consortium within the MII. As part of the ongoing Development for a research study monitor within the use case 1 of MIRACUM, the clinical study monitor tool has been configured to detect patients with rare diseases during their routine clinical encounters.

2. Methods and Results

To gain first experience with this concept, the study was limited to the ICD-10 code E84 for mucoviscidosis. Because UHG also operates a large pediatrics department and runs a special center for long term treatment of mucoviscidosis we were also able to compare visit frequencies for both the treatment center and other clinical departments.

To support the process of identifying patients with rare diseases, the MIRACUM clinical trial recruitment support system (CTRSS) [3] has been configured to identify patients with mucoviscidosis once they are readmitted for clinical treatment. The CTRSS is installed within the data context of the data integration center (DIC) at the UHG (

The CTRSS has been enhanced to deliver notifications to the patient data management system (PDMS) ICUdata on the intensive care units and activate a template for documentation of mucoviscidosis symptoms within the chart to enhance situational awareness for the disease symptoms and problems.

3. Discussion and Outlook

It could be shown that the CTRSS is able to detect patients with prior diagnosis of mucoviscidosis (ICD-10 E84) and could also decide whether a patient is treated on an intensive care ward.

Furthermore, it could be proven, that it is possible to place requests for further documentation automatically via HL7-messages from the CTRSS into the patient chart. According to the clinical guidelines for documentation within UHG wards, such requests must be fulfilled by the responsible clinicians within a working shift.

For further data analysis and system improvement a working group within the DIC consisting of a physician and a medical documentary has been established to analyze the resulting documentation and make adaptations to the documentation templates.

References

[1] Nationales Aktionsbündnis für Menschen mit Seltenen Erkrankungen (NAMSE) last accessed 10.1.2023: www.namse.de
[2] Use Case CORD-MI: https://www.medizininformatik-initiative.de/de/CORD
[3] Gulden C, Macho P, Reinecke I, Strantz C, Prokosch HU, Blasini R. recruIT: A Cloud-Native Clinical Trial Recruitment Support System based on HL7 FHIR and the OMOP CDM - Design and Implementation Study (Preprint). JMIR Prepr. 2022;23/08/2022(42139). doi:10.2196/preprints.42139.
[4] Blacketer C.: Chapter 4 The Common Data Model | The Book of OHDSI. OHDSI; 2019. Accessed August 24, 2022. https://ohdsi.github.io/TheBookOfOhds
[5] Michel A, Benson M, Junger A, Sciuk G, Hempelmann G, Dudeck J, Marquardt K.: Design principles of a clinical information system for intensive care units (ICUData). Stud Health Technol Inform. 2000;77:921-4. PMID: 11187689.
[6] https://www.medizininformatik-initiative.de/de/CORD: Use Case CORD-MI

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230140

Pharmacists' Experiences of Using Knowledge Bases on Medicines in Pregnancy and Breastfeeding

Linnéa KARLSSON LIND[a,1], Patrik DREHER SKÖLD[a] and Ulrika NÖRBY[a]
[a] *Health and Medical Care Administration, Region Stockholm, Sweden*
ORCiD ID: Linnéa Karlsson Lind https://orcid.org/0000-0002-9570-7783,
Ulrika Nörby https://orcid.org/0000-0003-0396-5986

Abstract. Knowledge bases on medicines during pregnancy and breastfeeding integrated into a clinical decision support system are valuable tools for pharmacists. The information facilitates counseling, is time-saving and improves patient safety.

Keywords. Clinical decision support system, pregnancy, breastfeeding, pharmacy

1. Introduction

Medicine treatment during pregnancy and breastfeeding implies special challenges. To facilitate access to reliable information, Region Stockholm provides the knowledge bases *Janusmed Drugs and Birth Defects* and *Janusmed Breastfeeding*, freely accessible on the internet (https://janusmed.se) and integrated into electronic health records. In September 2021, the knowledge bases were also integrated into the Electronic Expert Support (EES), a clinical decision support system used when dispensing prescription medicines, available at almost all Swedish pharmacies [1]. The EES is provided by the Swedish eHealth Agency and financed by the government. This study is a preliminary evaluation of the pharmacists' perceptions of using the two knowledge bases via EES.

2. Methods

A web-based questionnaire was distributed via e-mail or intranet six months after the launch of the databases in EES, to approximately 5 000 pharmacists in Swedish pharmacies. The questions focused on use and perceptions of the databases and were multiple-choiced, scaled, or open-ended. A survey tool provided by Questback Essentials® was used to set up and collect the data.

[1] Corresponding Author: Linnéa Karlsson Lind, E-mail: linnea.karlsson-lind@regionstockholm.se.

3. Results

In total, 403 pharmacists responded to the questionnaire (response rate ~ 8-10%). Respondents from all regions in Sweden participated, most of them (95%) worked at community pharmacies. The majority (77%) were aware of the two new knowledge bases in EES, and more than half had used them. The respondents perceived that the information is trustworthy, time-saving and improves patient safety (Fig.1).

Among the respondents, around 38% received questions every week from pharmacy customers regarding medicines in pregnancy or during breastfeeding. When asked how they handle these questions, the majority said they use Physicians' Desk Reference (94%). Many respondents also used EES (66%) and the website for the knowledge bases (59%).

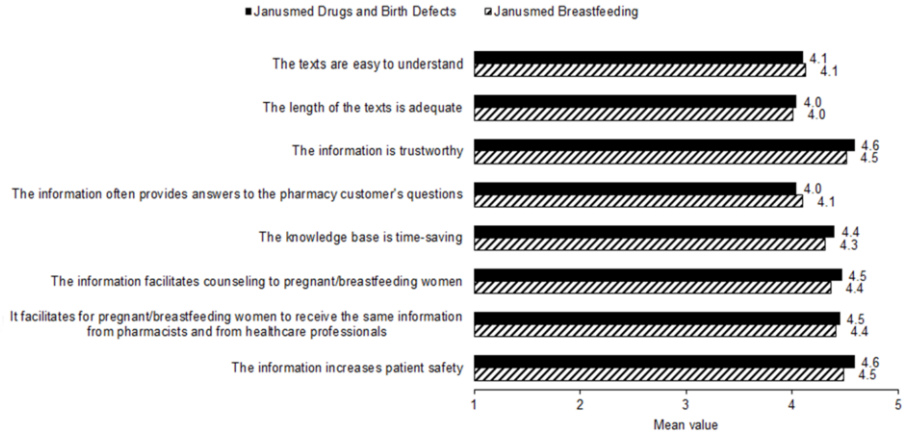


Figure 1. Responses to statements about *Janusmed Drugs and Birth Defects* and *Janusmed Breastfeeding* in Electronic Expert Support (EES), on a scale 1-5 (1=Totally disagree, 5=Totally agree).

4. Discussion and Conclusions

This study shows that pharmacists rate the two knowledge bases as reliable and supportive when dispensing prescription medicines. The positive results are in line with previous evaluations of the databases among health care professionals and pregnant women [2]. If health care professionals and pharmacists use the same information source, this most likely facilitates communication and ensures that pregnant and breastfeeding women receives consistent, professional advice.

References

[1] Hammar T, Hellström L, Ericson L. The Use of a Decision Support System in Swedish Pharmacies to Identify Potential Drug-Related Problems—Effects of a National Intervention Focused on Reviewing Elderly Patients' Prescriptions. Pharmacy (Basel). 2020 Jul 13;8(3):118. doi: 10.3390/pharmacy8030118

[2] Norby U, Kallen K, Shemeikka T, Korkmaz S, and Winbladh B. Pregnant women's view on the Swedish internet resource Drugs and Birth Defects intended for health care professionals. Acta Obstet Gynecol Scand. 2015;94:960-68. DOI: 10.1111/aogs.12702

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230141

Local Approval Processes in a Federated and Distributed Research Infrastructure - Lessons Learned from the AKTIN-Project

Alexander KOMBEIZ[a,1], Jonas BIENZEISLER[a], Saskia EHRENTREICH[b], Rainer RÖHRIG[a] and Raphael W. MAJEED[a] on behalf of the AKTIN Research Group
[a] *Institute of Medical Informatics, Medical Faculty of RWTH Aachen, Aachen, Germany*
[b] *Department of Trauma Surgery, Otto von Guericke University, Magdeburg, Germany*

Abstract. The AKTIN-Emergency Department Registry is a federated and distributed health data network which uses a two-step process for local approval of received data queries and result transmission. For currently establishing distributed research infrastructures, we present our lessons learned from 5 years of established operations.

Keywords. information networks, distributed systems, health services research, data sharing, routinely collected health data, electronic health records, AKTIN

1. Introduction

The AKTIN-Emergency Department (ED) Registry [1,2] is a federated and distributed research infrastructure providing access to standardized ED routine documentation [3]. Data collected during clinical routine are automatically stored in local data warehouses (DWH) of the participating institutions and can be queried using the central AKTIN Broker. When querying data from the local DWHs, each ED is responsible to review query requests for compliance with local ethical, legal and organizational requirements. The responsibility to review data requests is typically delegated. The responsible person needs all the information required to check conformity of the request. To allow for an informed decision, a data approval process was established. The AKTIN infrastructure has been in operation since 2017. As of Dec 8th, 2022, 48 participating clinics are connected. The objective of this work is to present the lessons learned from 5 years of operation.

The data approval process is integrated into the graphical user interface of the DWH. It is implemented as a two-step process (approval of data request and transmission of results to the AKTIN Broker) in a web application. The user receives a cover letter describing the purpose and content of the request. The user can check the cover letter using the attached metadata and technical syntax. Without the user's consent, a request cannot be executed, and the results cannot be transferred to the Broker. The user has the option to review the results before they are transmitted. For convenience, the user has the option to give a revocable consent of automatic approval or rejection of repeating requests.

[1] Corresponding Author: Alexander Kombeiz, E-Mail: akombeiz@ukaachen.de

2. Methods

We used descriptive statistics to describe the transmitted status changes of all requests stored on the AKTIN Broker and submitted between May 1st 2020 and Dec 7th 2022. Frequencies and percentages of approval, rejection and failure states as well as mean, median and interquartile range of the review time were calculated. We gathered insights from our operational experience and feedback from ED correspondents.

3. Results

During the 31-month period, a total of 470 data requests were sent, resulting in 10,787 interactions with the release process. In 80.5% (n=8680) of the interactions, the request was approved and the results transmitted, in 1.3% (n=141) the request was rejected, in 0.4% (n=44) the request failed during processing or result transmission and in 17.8% (n=1922) the request remained unassessed. 82.2% (n=7138) of all approval and 32.6% (n=46) of all rejection interactions occurred with automatic consent. The mean processing time between receiving a request and non-automatic approval was 19.29±37.42 days, median was 5.04 (IQR = 0.86-21.29) days. Non-automatic rejection took 44.79±53.33 days on average, median was 26.29 (IQR = 15-41.04) days. There was no known incident where a request was approved by mistake. In individual cases, questions of understanding were asked about the approval process.

4. Discussion and Conclusion

Fulfilling its purpose, the local approval process in the AKTIN registry was used extensively and without adverse events. The majority of requests were approved. We attribute the unevaluated interactions to the high workload within the institutions. It was helpful to track ongoing requests and notify correspondents of pending assessments. Automated email notifications were rarely reacted to. A median of 5 days to assess and approve a request is relatively quick, given the high-stress environment of EDs. The large IQR regarding the processing time of a request reflects a core group of hospitals that assess the requests promptly and reliably. As only timestamps are recorded, there is no information on the actual time spent on the assessment. The high proportion of automated approvals suggests a general trust in the procedure.

References

[1] Ahlbrandt J, Brammen D, Majeed RW, Lefering R, et al. Balancing the need for big data and patient data privacy-an IT infrastructure for a decentralized emergency care research database. Stud Health Technol Inform. 2014;205:750–4, doi: 10.3233/978-1-61499-432-9-750

[2] Brammen D, Greiner F, Kulla M, Otto R, et al. AKTIN – The German Emergency Department Data Registry – real-time data from emergency medicine. Med Klin Intensivmed Notfmed. 2020 Dec 21;1–10, doi: 10.1007/s00063-020-00764-2

[3] Kulla M, Baacke M, Schöpke T, Walcher F, et al. Core Dataset "Emergency Department" of the German Interdisciplinary Association of Critical Care an Emergency Medicin (DIVI). Notfall Rettungsmed. 2014 Dec 1;17(8):671–81, doi: 10.1007/s10049-014-1860-9

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230142

Deep Learning Method for Estimation of Morphological Parameters Based on CT Scans

Rajarajeswari GANESAN[a,1], Antonino A. La MATTINA [b,c], Frans N. van de VOSSE[a] and Wouter HUBERTS[a,d]
[a]*Eindhoven University of Technology, The Netherlands*
[b]*Alma Mater Studiorum - University of Bologna, Italy*
[c]*IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy*
[d]*Maastricht University, The Netherlands*

Abstract. In this study, we propose a Convolutional Neural Network (CNN) with an assembly of non-linear fully connected layers for estimating body height and weight using a limited amount of data. This method can predict the parameters within acceptable clinical limits for most of the cases even when trained with limited data.

Keywords. Height, Weight, Deep Learning, Electronic Health Records (EHRs), CNN.

1. Introduction

In this study, we investigate the performance of a CNN-based regression model to estimate height and weight using limited data (Femur scans of less than 50 patients). This Deep Learning method can extract features from images and potentially establish a relation between features and morphological parameters [1]. Estimation of height and weight using CT scans can be used for completing associated missing Electronic Healthcare Records, and/or generating synthetic patient data.

2. Method

Fig.1. shows a CNN model with convolution blocks and a stack of non-linear fully connected (FC) layers to map the hierarchical image features and the morphological parameters. In the assembly, the first FC layer is fed with the feature vector and output is as $Y_i^{(L_i)} = \phi\left(\sum_{p=1}^{f_1^{(c)}}\sum_{q=1}^{f_2^{(c)}}\sum_{r=1}^{f_3^{(c)}} w_{g,p,q,r}\left(P_p^{(c)}\right)_{q,r}\right)$ in which L_1 is the first dense layer, $f_1^{(c)}$ is the number of feature maps of size $f_2^{(c)} \times f_3^{(c)}$, $P^{(c)}$ is the output from the c[th] convolutional block, w is the weighted connection from position (q,r) in feature map p to g^{th} unit in layer L_1 and ϕ is the Leaky ReLU function [2]. The consequent FC layers act like Multilayer Perceptron with previous layer inputs. We have used limited training data (36

[1] Corresponding Author: Rajarajeswari Ganesan, Department of Biomedical Engineering, Eindhoven University of Technology, The Netherlands; E-mail: r.ganesan@tue.nl

patients/10k images) and two test sets (11 and 12 patients). The methods have been implemented using TensorFlow 2.9.0 Library and assessed using metrics: Absolute Error=$|y_i - \hat{y}_i|; 1 \le i \le S$; Percentage Accuracy=$\frac{N_{S<Thr}}{N_S} \times 100$ where $N_{S<Thr}$ is the number of patients with predictions less than or equal to 5 kg/cm (for more precise estimations) and N_S is the total number of patients where S is the size of the test dataset.

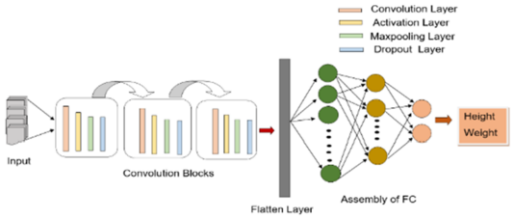


Figure 1. Architecture of Stacked-CNN Model.

3. Results and Discussion

Table 1 shows the predictions of height and weight for two patients from both test sets by Stacked-CNN. It also shows that the overall accuracy of Stacked-CNN is larger than 63% for estimating height in both sets, whereas the weight estimations are less good in Set 2 (42%). However, absolute differences between measurements and the actual value are within clinically acceptable limits (respectively , ±10 kg or ±10 cm) for all patients.

Table 1. Estimations and assessment of the method proposed [H: Height (in cm); W: Weight (in kg)]

Test Set ID	Actual(H /W)	Predicted (H/W)	Accuracy for (H/W)	Absolute Error for (H/W)
Set 1	159/60	158.6/60.1	63.6% (H)	4.5±2.8 (cm)
Set 1	150/50	156.5/54.6	63.6% (W)	5.0±3.3 (kg)
Set 2	164/65	159.9/57.8	75% (H)	3.1±2.0 (cm)
Set 2	158/58	162.5/60.3	41.7% (W)	5.4±2.8 (kg)

4. Conclusion and Future Work

The Stacked-CNN model shows good accuracy for the threshold of deviation (±5 kg or ±5 cm **:** acceptable limit for meticulous estimations) for body weight/height respectively. Due to relationship class imbalance, i.e., inequal number of patient datasets for the height-weight relation, height is predicted more accurately than weight.

Acknowledgement: The authors acknowledge the EU funded project In Silico World, grant number 101016503. The dataset is available at http://doi.org/10.6092/unibo/amsacta/7126 under the terms of the CC BY 4.0 International license.

References

[1] Ichikawa S, Hamada M, Sugimori H. A deep-learning method using computed tomography scout images for estimating patient body weight. Scientific reports.2021;11(1):1-9

[2] Stutz D. Understanding Convolutional Neural Networks. Seminar, Fakultät für Mathematik, Informatik und Naturwissenschaften Lehr- und Forschungsgebiet Informatik VIII Computer Vision; 2014.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230143

Towards an Advanced Digital Infrastructure Within the Non-University Sector Demonstrated by the PICOS App

Dr. Denise MOLINNUS[a,1], Angelique KURTH[a], Volker LOWITSCH[b], Prof. Gernot MARX, FRCA[a] and Prof. Johannes BICKENBACH[a]
[a] *Operative Intensivmedizin und Intermediate Care, Universitätsklinikum RWTH, Aachen, Germany*
[b] *Healthcare IT Solutions GmbH, Germany*
ORCiD ID: Dr. Denise Molinnus https://orcid.org/0000-0002-2140-259X

Abstract. The non-university sector is a central facility for the medical care of patients in Germany. So far, information technology infrastructure in this local health care sector is not developed and the many generated patient data are not further used. In this project, an advanced integrative, digital infrastructure will be established within the regional health care provider. Furthermore, a clinical use case will demonstrate the functionality and added outcome value of cross-sectoral data with a newly developed app to support follow-up care of former intensive care unit patients. The app will give an overview of current health status and generate longitudinal data for use in further clinical research.

Keywords. ICU, PICS, app development, longitudinal data, interoperability

1. Introduction

Large amounts of medical data are generated in acute care facilities, but also in downstream structures like rehab centers, nursing care or medical practices. However, only little innovation exists interlinking these chains of patient care and thus generating one available, longitudinal set of patient data. Although first approaches might be applied in high care facilities like university hospitals, particularly non-university areas lack of such developments. Within the framework of the progress hub DISTANCE (Digital Smart Hub for Advanced Connected Care), the interoperable data exchange is extended to medical facilities of regional care. A clear benefit for patients, but also for medical professionals and science is demonstrated in a clinical use case designing a special, former ICU-patient oriented application. In comparison to other existing patient-oriented apps such as ThessHF [1] or Ada [2] the use case "Post Intensive Care Outcome Surveillance" (PICOS) is aimed at capturing cross-sector longitudinally patient data at different times after treatment on ICU. The data are collected from patients prone to suffer from psychological and physical complaints known as Post Intensive Care Syndrome (PICS) [3] after prolonged intensive care treatment. In the long term, resulting

[1] Corresponding Author: Dr. Denise MOLINNUS, E-mail: dmolinnus@ukaachen.de.

data collected via the app will be anonymized, merged and made available for secondary data analysis and research purposes.

2. Methods

The overall task of the progress hub is to incorporate the pioneering work of the German medical informatics initiative on digitalization in medicine from university hospitals into the non-university health care system. Within the use case PICOS, a patient-oriented app is developed. For the first time, it will be possible to collect a variety of vital parameters as longitudinal data from the patients' homes while documenting psychological and physical conditions of ICU outpatients, that are susceptible for the PICS. After a prolonged ICU stay, while still in the hospital, health care personnel will approach fitting patients and engage and teach them in usage of the app. For controlling and improving of their functional outcome, the PICOS app will collect information on the individual patient state of health, and it will also support medical self-care in daily life. Via this app, patients will have a constant overview over their health status through answering a daily or weekly questionnaire. This questionnaire includes among other things vital signs, sleep quality and pain assessment, all indicators of the PICS [3]. Moreover medication, medical appointments and documents can be saved in one place. As a further motivation, after periods 1, 3, 6 and 12 months, patients using the app will be invited to participate in free follow up examinations. For the first time, we present the PICOS app which bundles information on the mental and physical condition, while large amounts of data are generated with which the data situation on the long-term course of critical illnesses can be improved.

3. Discussion and Conclusion

The DISTANCE project will be using methods of data management and data modeling. Beside merging all incurring data of such a complex and prolonged treated patient group within different levels of care, previously unknown relationships and dependencies of various diseases might be identified. Clinicians, researchers, computer scientists and other health professionals can use the collected data to optimize clinical conditions and treatment processes for specific patient groups by providing and analyzing long-term observational data using artificial intelligence methods.

References

[1] Bakogiannis C, Tsarouchas A, Mouselimis D, Lazaridis C, Theofillogianakos E, Billis A, Tzikas S, Fragakis N, Bamidis P, Papadopoulos C, Vassilikos V. A Patient-oriented App (ThessHF) to improve self-care quality in heart failure: From evidence-based design to pilot study, JMIR Mhealth Uhealth 2021;9(4):e24271, URL: https://mhealth.jmir.org/2021/4/e24271, doi: 10.2196/24271.

[2] Wattanapisit A, Teo CH, Wattanapisit S, Teoh E, Woo WJ, Ng CJ. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks. BMC Med Inform Decis Mak. 2020 Jan 6;20(1):5. doi: 10.1186/s12911-019-1016-4. PMID: 31906985; PMCID: PMC6945711.

[3] Ramnarain D, Aupers E, den Oudsten B, Oldenbeuving A, de Vries J, Pouwels S. Post Intensive Care Syndrome (PICS): an overview of the definition, etiology, risk factors, and possible counseling and treatment strategies. Expert Rev Neurother. 2021 Oct;21(10):1159-1177. doi: 10.1080/14737175.2021.1981289. Epub 2021 Sep 22. PMID: 34519235.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230144

odML-Tables as a Metadata Standard in Microneurography

Alina TROGLIO[a,1], Fabian SCHLEBUSCH[a], Rainer RÖHRIG[b], James DUNHAM[c], Barbara NAMER[a, 2] and Ekaterina KUTAFINA[b, 2]
[a] *Junior Research Group Neuroscience, Interdisciplinary Center for Clinical Research Within the Faculty of Medicine, RWTH Aachen University, Aachen, Germany*
[b] *Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[c] *School of Physiology, Pharmacology and Neuroscience Bristol Anaesthesia, Critical Care and Pain Research, University of Bristol, UK*
ORCiD ID: Alina Troglio https://orcid.org/0000-0001-7770-4389

Abstract. Metadata standards are well-established for many types of electrophysiological methods but are still lacking for microneurographic recordings of peripheral sensory nerve fibers in humans. Finding a solution for daily work in the laboratory is a complex process. We have designed templates based on odML and odML-tables to structure and capture metadata and provided an extension to the existing GUI to enable database searching.

Keywords. Metadata, FAIR, open metadata Markup Language (odML), microneurography, electrophysiology, pain

1. Introduction

Metadata standards are essential for compliance with FAIR principles and collaborative work between research sites. Microneurography is an electrophysiological method for observing the activity of single peripheral nerve fibers in awake healthy humans and patients with neuropathy, chronic pain or itch [1]. Multicenter collaborations in microneurography are particularly important due to relatively small numbers of recordings being made by individual labs, making data stratification challenging. Currently, there are no standards for managing metadata of recordings. We present an open-source solution based on odML-tables [2,3], which provides an organized and standardized way to build a metadata database and offers a novel extension to the odML-tables GUI with search capabilities that increases the efficiency of daily laboratory routines.

2. Requirements

We established the metadata handling requirements by several researchers from microneurography labs in Aachen and Bristol based on their experience with the

[1] Corresponding Author: Alina Troglio, E-mail: alina.troglio@rwth-aachen.de
[2] Both authors contributed equally to the paper.

experimental setup and data. The solution needs to be simple to use and easily adaptable for a large variety of experimental setups. Additionally, the data has to be stored separately from the metadata as the metadata contains sensitive clinical information. Only fully anonymized raw data can be shared between labs after relevant filtering (e.g., stimulation protocol, fiber types, patient diagnoses) is done internally.

3. Results

We have designed two templates in a human-readable table format that can be converted to odML format. One template contains general experimental information, and the other template holds details about the recordings themselves. In the Aachen lab, the templates are filled in for the most recent experiments and aggregated into a single odML database. A search function has been added to the existing odML-tables GUI to increase the lab routine productivity and simplify the retrieval of recordings for specific research questions. The templates, instructions, and odML-tables modifications can be found in our GitHub repository[3].

4. Discussions and Conclusions

The creation of a metadata database and the implementation of the search functionality are already supporting the routine work of the Aachen microneurography lab and we expect full functionality to be available in the Bristol lab in the next months. In the next steps, we aim to discuss our solution with other representatives of the microneurography community. Further, we are working on the computational pipeline openMNGlab [4], which integrates different raw data formats of microneurography through Neo [5] structure and supports further analytical processes. Automatizing metadata search and subsequent loading of the raw data (possible only on internal lab data) is an important step to improving the efficiency of microneurography computing and thus getting a better insight into the processes of peripheral signaling of pain and itch.

References

[1] Ackerley R, Watkins RH. Microneurography as a tool to study the function of individual C-fiber afferents in humans: responses from nociceptors, thermoreceptors, and mechanoreceptors. J Neurophysiol. 2018 Dec 1;120(6):2834–46.

[2] Grewe J, Wachtler T, Benda J. A bottom-up approach to data annotation in neurophysiology. Front Neuroinformatics. 2011;1–18.

[3] Sprenger J, Zehl L, Pick J, Sonntag M, Grewe J, Wachtler T, et al. odMLtables: A User-Friendly Approach for Managing Metadata of Neurophysiological Experiments. Front Neuroinformatics 2019;13.

[4] Schlebusch F, Kehrein F, Röhrig R, Namer B, Kutafina E. openMNGlab: Data Analysis Framework for Microneurography - A Technical Report. Stud Health Technol Inform. 2021 Sep 21;283:165–71.

[5] Garcia S, Guarino D, Jaillet F, Jennings T, Pröpper R, Rautenberg P, et al. Neo: an object model for handling electrophysiology data in multiple formats. Front Neuroinformatics. 2014.

[3] https://github.com/Digital-C-Fiber/odMLtablesForMNG

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230145

Medical Apps for Android and iOS: Differences and Similarities

Evgenii PUSTOZEROV[a,1], Ute VON JAN[b], Dennis LAWIN[a,c] and Urs-Vito ALBRECHT[a]
[a] *Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Bielefeld, Germany*
[b] *Peter L Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
[c] *Department of Cardiology and Intensive Care Medicine, University Hospital OWL of Bielefeld University, Campus Klinikum Bielefeld, Bielefeld, Germany*
ORCiD ID: Evgenii Pustozerov https://orcid.org/0000-0001-8109-1319, Ute von Jan https://orcid.org/0000-0001-9225-593X, Dennis Lawin https://orcid.org/0000-0002-1292-7845, Urs-Vito Albrecht https://orcid.org/0000-0002-8989-6696

Abstract. Google Play and Apple's App Store dominate the mobile health app market. We analyzed the metadata and descriptive texts of apps in the medical category using semi-automated retrospective app store analysis (SARASA) and compared the store offerings in terms of their number, descriptive texts, user ratings, medical device status, diseases, and conditions (both keyword-based). Relatively speaking, the store listings for the selected items were comparable.

Keywords. mHealth, mobile apps, Google Play, App Store, mobile health

1. Introduction

Healthcare professionals need relevant information about the applications to be in a position to make informed usage decisions [1]. Previous analyses showed that manufacturers often only provide inadequate information via meta-data [2] or within the app description texts. This article examines the discoverability of health app descriptions and metadata for Google Play Store and Apple's App Store, highlighting similarities and differences between the two major providers of health and medical apps.

2. Methods

R- and Python-based scripts were utilized to obtain data for both Google Play and the Apple App Store. Data parsing, processing, and visualization were done in R (version 4.2.2) and Python (version 3.10). We applied data filters to eliminate incomplete app data and duplicates. The stores were compared concerning their number, descriptive texts,

[1] Corresponding Author: Evgenii Pustozerov, Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany; E-mail: evgenii.pustozerov@uni-bielefeld.de.

user ratings, medical device status, diseases, and conditions. Descriptive statistics: percentages, ranges, median values, and interquartile ranges (IQR) were calculated.

3. Results

For Apple's US App Store, from the metadata for the original lists of 35,098 apps collected between Dec. 05, 2022, and Dec. 10, 2022, we identified 25,682 apps with English language store descriptions. For Google Play, we were able to download data for 11,365 apps with English descriptions. The median of the average user ratings assigned to the apps in the medical category was 4.10 (IQR: 3.50, 4.50) for the Android platform and 4.56 (IQR: 3.44, 5.0) for Apple. Ratings were missing in 8,388 (74%) of the apps in Google Play and 9,484 (42%) of the apps in Apple's App Store.

There were 1,896 apps in Google Play and 3,744 apps in Apple's App Store that matched one or more of the terms: blood pressure, cancer, cardio, corona, covid, depression, diabetes, heart, infection, stroke, and tumor. The proportions of apps matching any of the keywords were largely similar (16.7% matched on Google Play, respectively 16.6% on Apple's App Store). However, there were notable differences regarding a subset of the keywords. We found more apps in the "cardio" (3.9% vs. 2.8%), "diabetes" (4.2% vs. 2.7%), and "infection" contexts (3.2% vs. 2.6%) in the Google Play Store. Apps matching "heart" were more common on Apple's platform (5.5% vs. 3.2%).

Assessing the medical device status showed that only a negligible proportion of apps provided information in a readily identifiable manner. For the disease-related apps matching any of the terms, there were 188/3,744 (5%) on Apple's App Store and 80/1,896 (4.2%) on Google play that mentioned any of the medical device-related terms (either acknowledging or negating device status) in the description.

4. Conclusion

Overall, the two major app repositories do not differ much in the samples we acquired, relatively speaking, regarding average user ratings, the average length of descriptive text, and medical device status information. Currently, there is no easy and cost-effective way to analyze the full range of a large app repository ad-hoc. There is also no defined app evaluation methodology according to recent reviews [3]. Healthcare professionals, who rely on valid information about the apps, are disadvantaged by this situation. It would be helpful if store operators were more transparent about their offers, for example, by establishing gateways or store listings reporting the necessary aspects.

References

[1] Albrecht UV, Framke T, von Jan U. Quality awareness and its influence on the evaluation of app meta-information by physicians: validation study. JMIR mHealth and uHealth. 2019 Nov 18;7(11):e16442.. doi:10.2196/16442.

[2] Pustozerov E, von Jan U, Albrecht UV. Evaluation of mHealth applications security based on application permissions. InUnifying the Applications and Foundations of Biomedical and Health Informatics 2016 (pp. 241-244). IOS Press.

[3] Alamoodi AH, et al. A systematic review into the assessment of medical apps: motivations, challenges, recommendations and methodological aspect. Health and Technology. 2020 Sep;10:1045-61.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230146

Implementation of Interoperable Healthcare Standards for Community Healthcare

Ye Seul BAE MD, MPH [a,b], Yeju PARK [a,1], Seung Min LEE [c], Hahyun YOU [c], Yujin PARK [c], Hyeonji LEE [a] and Hyung-Jin YOON [c]
[a]*Department of Family Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, South Korea*
[b]*Department of Future Healthcare Planning, Kangbuk Samsung Hospital, South Korea*
[c] *Department of biomedical engineering, College of Medicine, Seoul National University Hospital, Seoul, South Korea*

Abstract. Building an integrated data model that includes not only clinical data but also personal health records has become increasingly important. We aimed to build a big data healthcare platform by developing a common data model that can be utilized in the healthcare field. To this end, we acquired health data from various communities to establish community care digital healthcare service models. Further, to improve personal health data interoperability, we ensured conformance to international standards, namely, the Systemized Nomenclature of Medicine Clinical Terms (SNOMED-CT) and transmission standards, namely, Health Level 7 Fast Healthcare Interoperability Resource (HL7 FHIR). Furthermore, FHIR resource profiling was designed to transmit and receive data, following the HL7 FHIR R4 guidelines.

Keywords. Healthcare standards, patient generated health data, bigdata platform, Systemized Nomenclature of Medicine Clinical terms, HL7, FHIR

1. Introduction

In this study, we aimed to build an open big data platform for personal healthcare. We provided standards for establishing clinical terminologies, structuring data, data security, and quality management to improve data interoperability. This standardized system will facilitate data integration, management, and utilization.

2. Methods and Results

Standardization with SNOMED-CT and FHIR: We defined a common dataset for SNOMED-CT concepts that are frequently used by clinical institutions. A common dataset was created using data obtained from Korea's National Health Screening Program [1] and health Application Programming Interfaces (APIs) from global corporations, such as Samsung [2] and Google [3]. To ensure semantic interoperability via mapping clinical terms, we designed customized SNOMED post-expressions based on the FHIR structure (Figure 1). In addition, we profiled the FHIR, obtaining a transmission standard called HL7 FHIR, to set the data model.

[1] Corresponding Author: Yeju Park, E-mail: armeniaca04@gmail.com.

Evaluating the generality of standardized common data model: In the healthcare field, big data sources are obtained under various situations. Through FHIR resource profiling using the SNOMED-CT expression, clinical indicators in the data model can be expressed consistently without changing the model structure and representative observation code. To ensure consistency and accuracy when mapping post-coordinated code, we used the validation tool provided by SNOMED CT International and HL7 [4]. In addition, three experts in this field assessed validation results via cross-checking or group discussion [5].

Development of bigdata platform: After standardizing the data, we developed a big data platform to integrate data from various sources. Researchers, healthcare professionals, and healthcare companies will be able to utilize the platform. Pseudonymization processing is performed before transmitting data to the platform. Additionally, we follow a data security guideline that complies with Korea's Personal Information Protection Act, and we have developed a cloud-based platform. Furthermore, we plan to deploy a system-theoretical, architecture-centric, ontology-based, policy-driven approach that conforms to the ISO 23903 standard.

3. Discussion and Conclusions

The challenge associated with big data management was limited entity sets that cover the physical data used in clinical institutions. Expansion of the service model and domain in personal areas, such as psychological data, will facilitate the building of an optimized platform that enables individuals to proactively use and share their health information through a big data platform with improved interoperability.

To standardize the clinical terminology, we mapped clinical terminology codes using SNOMED-CT. In addition, to appropriately set the data models, we developed transmission standard guidelines by profiling the HL7 FHIR resource. We designed medical institution datasets to simplify the exchange of clinical information among disparate healthcare providers and health service systems. Furthermore, we expect that using international standards data and linking personal health data will facilitate healthcare research and healthcare service industry.

Acknowledgements: This research was supported by a grant of the Korea Health Promotion R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HS21C0016).

References

[1] Shin DW, Cho J, Park JH, Cho BL. National General Health Screening Program in Korea: history, current status, and future direction: A scoping review. Precision and Future Medicine 2022; 6(1): 9-31,

[2] Samsung [Internet]. c2020. Samsung Health Android SDK Data API Reference 1.5.0. Available at https://img-developer.samsung.com/onlinedocs/health/android/data/index.html [accessed on 3 Mar 23].

[3] Google [Internet]. c2021. Google Fit Guides. Available at https://developers.google.com/fit/datatypes [accessed on 3 Mar 23].

[4] SNOMED International, “Expression Constraint Language - Specification and Guide - Expression Constraint Language - SNOMED Confluence (ihtsdotools.org)”

[5] Kang H, Park HA. Mapping Korean National Health Insurance Reimbursement Claim Codes for Therapeutic and Surgical Procedures to SNOMED-CT to Facilitate Data Reuse. Stud Health Technol Inform. 2022 Jun 6;290:101-105. doi: 10.3233/SHTI220040.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230147

Status of Phlebitis in South Korean Hospitals: Focusing on Electronic Incident Reporting Systems

Insook CHO[a], Jae-Ho LEE[b] and Eun Man KIM[c,1]
[a] *Department of Nursing, Inha University, Incheon, South Korea*
[b] *Department of Information Medicine, Department of Emergency Medicine, Asan Medical Center, Seoul, South Korea*
[c] *Department of Nursing Science, Sun Moon University Chungnam, South Korea*
ORCiD ID: Insook Cho https://orcid.org/0000-0002-5152-9567, Jae-Ho Lee https://orcid.org/0000-0003-2619-1231, Eun Man Kim https://orcid.org/0000-0002-9470-8936

Abstract. Peripheral venous catheterization (PVC) is the most commonly used invasive technique, and its importance to patient safety is increasing. And phlebitis is a common complication which can lead to increased costs and extended hospital stays. This study attempted to characterize the current status of phlebitis based on incident reports in the Korea Patient Safety Reporting & Learning System. This retrospective descriptive study analysed 259 phlebitis cases reported in that system from 1 July 2017 to 31 December 2019. The analysis results were summarized using numbers and percentages or means with standard deviations. Among the reported phlebitis cases, antibiotics and high-osmolarity fluids comprised 48.2% of the intravenous inflammatory drugs used. All reported cases presented blood-flow infections. Insufficient observation or management was the most common cause of phlebitis. It was found that interventions for phlebitis were inconsistent with those recommended in evidence-based guidelines. Recommendations for nurses to alleviate complications in PVC must be promoted and educated. It is necessary to provide feedback from the incident reports analysis.

Keywords. Phlebitis; Patient safety, Incident report

1. Introduction

Peripheral venous catheterization (PVC) is the most commonly used invasive medical technique in hospitals, and its importance to patient safety is increasing [1,2]. Phlebitis is a common complication associated with peripheral catheter use, which can lead to increased treatment costs and extended hospital stays [3]. Reducing incidents that endanger patient safety has become a major concern in healthcare today [3]. To achieve the goal of reducing medical errors, electronic incident reporting systems should be secure, easy to use and effective [4]. The Patient Safety Act was implemented in South Korea in July 2016. The enforcement of the associated protocols resulted in the

[1] Corresponding Author: Eun Man KIM, E-mail: kem7355@hanmail.net.

development of a nation-level electronic reporting systems, with hospitals nationwide reporting patient safety incidents to this system. This study aimed to identify intravenous factors related to the occurrence, causes, post-discovery responses of phlebitis based on reports from electronic reporting systems in South Korea.

2. Methods

This retrospective descriptive study aimed at characterizing the status of phlebitis in South Korea. 259 phlebitis cases from hospitals throughout the country were selected from 1 July 2017 to 31 December 2019. The findings are summarized using numbers and percentages or mean±standard-deviation values.

3. Results

Among the reported phlebitis cases, 44.5% patients had inflammatory disease or tumours. and 25% of patients were infused with a hyperosmotic solution. All of the cases were found to have suffered from blood-flow infections. Insufficient observation or management was the most common cause of phlebitis. Cold packs were used as an intervention for phlebitis in the largest proportion of cases, followed by catheter removal.

4. Discussion and Conclusions

Awareness of appropriate PVC maintenance among nurses and their early identification of the risk factors can help to minimize potential adverse results. Providing nurses with training and support on recommendations for alleviating complications in PVC management is therefore required. These electronic reporting systems lack analysis and feedback on collected incident reports. It is necessary to further develop such learning system. Limitation is the retrospective analyses of phlebitis-related reports, which resulted in the study not being able to identify causal relationships between the risk factors and the occurrence.

The findings of this study support the periodic assessment of IV sites, application of aseptic procedures and compliance with standards. It is necessary to develop and apply educational programs about PVC management and infection control for nurses. Moreover, a learning system is needed for analysing incident reports and providing feedback from them.

References

[1] Nickel B. Peripheral intravenous administration of high-risk infusions in critical care: a risk-benefit analysis. Crit Care Nurse 2019;39(6):16–28. doi: 10.4037/ccn2019443.
[2] Tan YHG, Tai WLS, Sim C, Ng HLI. Optimising peripheral venous catheter usage in the general inpatient ward: a prospective observational study. J Clin Nurs 2017;26(1–2):133–9. doi: 10.1111/jocn.13451.
[3] González-López JL, et al Portal P. Indwell times, complications and costs of open vs closed safety peripheral intravenous catheters: a randomized study. J Hosp Infect 2014;86(2):117–26.
[4] Leape LL. Reporting of medical errors: time for a reality check. Qual Health Care. 2000;9(3):144–5.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230148

Towards the Definition of an Intelligent System for Organizing Medical Visits and Collecting Medical Data

Evangelos LOGARAS[a,1], Antonis BILLIS [a], Evangelos STAMKOPOULOS [a], Paraskevas LAGAKIS [b], Ilias DIMITRIADIS [c], Athena VAKALI [c] and Panagiotis D. BAMIDIS [a]
[a] *Laboratory of Medical Physics and Digital Innovation, School of Medicine, Aristotle University of Thessaloniki, Greece*
[b] *Computer Center Lagakis, Greece*
[c] *Data and Web Science Laboratory, Department of Informatics, Aristotle University of Thessaloniki, Greece*
ORCiD ID: Evangelos Logaras https://orcid.org/0000-0003-4921-3819
ORCiD ID: Antonis Billis https://orcid.org/0000-0002-1854-7560
ORCiD ID: Evangelos Stamkopoulos https://orcid.org/0000-0002-0646-1364
ORCiD ID: Paraskevas Lagakis https://orcid.org/0000-0002-0605-9172
ORCiD ID: Ilias Dimitriadis https://orcid.org/0000-0003-1336-6960
ORCiD ID: Athena Vakali https://orcid.org/0000-0002-0666-6984
ORCiD ID: Panagiotis Bamidis https://orcid.org/0000-0002-9936-5805

Abstract. Appointment Scheduling (AS), typically serves as the basis for the majority of non-urgent healthcare services and is a fundamental healthcare-related procedure which, if done correctly and effectively, can lead to significant benefits for the healthcare facility. The main objective of this work is to present ClinApp, an intelligent system able to schedule and manage medical appointments and collect medical data directly from patients.

Keywords. Medical Appointment Scheduling, Medical Appointment Management, Medical Data Management, Chatbot, Medical Questionnaire, Conversational Agent

1. Introduction

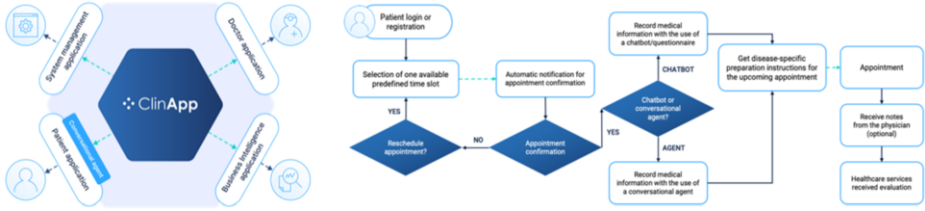
Medical appointments are traditionally scheduled by the secretary of the respective department or clinic, mainly via telephone communication. This method, although fully flexible, is directly affected by the human factor [1]. Digital healthcare is constantly evolving and the recent pandemic has reinforced the need to transition to more patient-centric approaches [2]. Thus, incorporating technology into the AS process through the use of web-based interventions, addressed directly to patients, may have a positive impact. To this end, this paper proposes the design of an intelligent Web-based Appointment System (WAS) for AS and remote medical data collection, called ClinApp.

[1] Corresponding Author: Evangelos Logaras, Laboratory of Medical Physics and Digital Innovation, Medical School, Aristotle University of Thessaloniki, Greece; E-mail: evanlogar@gmail.com.

2. Methodology and Results

In the context of this study, input from medical and administrative personnel was gathered through targeted interviews in order to determine the functional requirements needed for an efficient AS. The interviews resulted in 16 functional requirements that were used to design the ClinApp system.

The novelty of the proposed approach lies in the fact that besides AS, patients will be able to effortlessly and securely record their medical data prior to their appointment, thus gaining more quality time with the physician. In addition, patients will be able, if they want, to record their medical data through speech recognition using a conversational agent, while after confirming the completion of the appointment they will be given the opportunity to evaluate the services they received, providing that way valuable information to the administrative staff of the healthcare facility. For the reader's convenience, in Figure 1 the proposed system's (a) main applications and (b) AS procedure workflow are provided.



(a) (b)

Figure 1. ClinApp system's: a) Main applications and b) Workflow of the AS procedure.

3. Discussion and Conclusions

This work presents an initial attempt towards the definition of an intelligent system able to schedule and manage medical appointments and collect medical data directly from patients. Future steps include the technical development of the ClinApp system and then piloting it in private practices and hospital outpatient departments, where its overall performance and perceived usefulness by end users will be evaluated.

Acknowledgements

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (T2EDK-04937).

References

[1] Zhang X, Yu P, Yan J. Patients' adoption of the e-appointment scheduling service: A case study in primary healthcare. Stud Health Technol Inform. 2014;204:176-81. PMID: 25087546.

[2] Tebeje T, Klein J. Applications of e-Health to Support Person-Centered Health Care at the Time of COVID-19 Pandemic, Telemedicine and e-Health. 2021;150-158, doi: 10.1089/tmj.2020.0201

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230149

A Framework for Evaluating Synthetic Electronic Health Records

Emmanuella BUDU[a,1], Amira SOLIMAN[a], Kobra ETMINANI[a] and Thorsteinn RÖGNVALDSSON[a]
[a]*Center for Applied Intelligent Systems Research, Halmstad University, Sweden*
ORCiD ID: Emmanuella Budu https://orcid.org/0000-0003-4221-5467

Abstract. Synthetic data generation can be applied to Electronic Health Records (EHRs) to obtain synthetic versions that do not compromise patients' privacy. However, the proliferation of synthetic data generation techniques has led to the introduction of a wide variety of methods for evaluating the quality of generated data. This makes the task of evaluating generated data from different models challenging as there is no consensus on the methods used. Hence the need for standard ways of evaluating the generated data. In addition, the available methods do not assess whether dependencies between different variables are maintained in the synthetic data. Furthermore, synthetic time series EHRs (patient encounters) are not well investigated, as the available methods do not consider the temporality of patient encounters. In this work, we present an overview of evaluation methods and propose an evaluation framework to guide the evaluation of synthetic EHRs.

Keywords. Synthetic data, Electronic Health Records, evaluation

1. Introduction

Synthetic data should maintain the statistical and structural properties of real data without compromising the privacy of the individuals in the real data. Three main criteria: *fidelity*, *utility*, and *privacy*, are used to assess the quality of generated data [1,2]. The *utility* determines the usefulness of synthetic data for predictive and modelling purposes. *Fidelity* assesses the faithfulness of the synthetic data to real data. Finally, *privacy* assesses whether the privacy of the real data is compromised in the synthetic data.

However, several challenges exist. Several comparison methods and measures have been used, and new measures are often introduced in publications. This makes it difficult to compare the data generated by different models as there is no consensus on how to evaluate and compare the synthetic data generated by different models [2]. Secondly, the available methods do not evaluate the variable dependencies or consider the temporality found in patient encounters. They only focus on assessing synthetic EHRs as frozen in time, without dependencies between subsequent entries for the same individuals.

[1] Corresponding Author: Emmanuella Budu, E-mail: emmanuella.budu@hh.se.

2. Methods

In our approach, we first review the available methods and devise a hierarchy of evaluation methods categorized according to the type of data (e.g., categorical, continuous, discrete) and the mode of application (e.g., patient level, cohort level, and feature level). We aim to develop an evaluation framework guided by this hierarchy as illustrated in Figure 1 to assess the quality of synthetic EHRs.

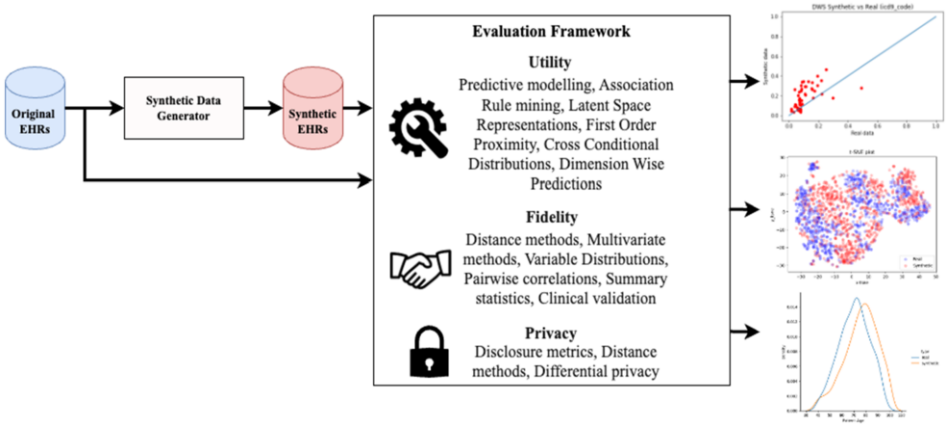


Figure 1. Proposed evaluation framework showing the original and synthetic EHRs passed as input to the evaluation framework to obtain qualitative and quantitative assessments. For example, the plots show the distribution of variables in the original and synthetic EHRs.

3. Results, Discussion, and Conclusions

Our categorization of existing evaluation methods identified several different methods of assessment under fidelity, utility and privacy [1,2]. Under fidelity: distance-based methods, variable distribution methods, correlations, comparison of data statistics, clinical validations, and multivariate methods. For utility: predictive modelling, dimension-wise predictions, association rule mining, first-order proximity, latent space representations, and cross-conditional distributions. For privacy: we have disclosure metrics, and distance-based methods.

This work mainly presents an overview of evaluation methods for synthetic EHRs. We propose a framework to standardize the evaluation of synthetic EHRs. We aim to make the implementation of the evaluation framework publicly accessible for use by other researchers in future.

References

[1] Mendelevitch O, Lesh M. Fidelity and Privacy of Synthetic Medical Data, Review of Methods and Experimental Results. ArXiv 2021:39. https://doi.org/10.48550/arXiv.2101.08658.

[2] Hernandez M, Epelde G, Alberdi A, Cilla R, Rankin D. Synthetic data generation for tabular health records: A systematic review. Neurocomputing 2022;493:28–45. https://doi.org/10.1016/j.neucom.2022.04.053.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230150

Protocol of Validation Mechanisms Within the Design of an Electronic Case Report Form for the MULTI-SITA Project

Gabriele DI MECO[a,1], Daniele Roberto GIACOBBE[a,b], Sara MORA[c], Silvia DETTORI[a,b], Matteo BASSETTI[a,b] and Mauro GIACOMINI[c]; on behalf of SITA (Italian Society of Anti-Infective Therapy)
[a] *IRCCS San Martino Polyclinic Hospital, Genoa, Italy*
[b] *Department of Health Sciences, University of Genoa, Genoa, Italy*
[c] *Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy*

Abstract. With the wide diffusion of web technology, dedicated electronic Case Report Forms (eCRFs) became the main tool for collecting patient data. The focus of this work is to thoroughly consider the data quality in every aspect of the design of the eCRF, with the result of having multiple steps of validation that should produce a diligent and multidisciplinary approach towards every step of data acquisition. This goal affects every aspect of the system design.

Keywords. eCRF, interface design, data quality, validation protocol

1. Introduction

Gathering data is central to the health care process [1], but it comes with the cost of a general complexity, that affects every aspect of the systems used to collect and validate the data [2]. Electronic Case Report Forms (eCRFs) have already proven themselves to be the key tools to achieve the control over the difficulties that the need of large and complete data provokes, however not every design guarantees the same disturbance rejection and completeness of data quality [3]. A strict collaboration in the development of our eCRF with the medical researchers of the Italian Society of Anti-Infective Therapy (*Società Italiana di Terapia Anti-Infettiva, SITA)* constituted a key aspect in the development our common platform (MULTI-SITA project) [4][5].

2. Materials and Methods

The web application is constituted by a client, written in *Blazor*, a Microsoft SQL Server database and an *API RESTful service*, which handles the communications between them. We followed acknowledged design principles for creating an ergonomic GUI [2][3].

[1] Corresponding Author, Gabriele Di Meco, IRCCS San Martino Polyclinic Hospital; E-mail: gabrieledimeco@yahoo.it.

3. Results and Discussion

The Database schema consists in a central entity – the patient – with multiple related forms that represent different aspects of the hospitalization, such as Medical History, Microbiology and Laboratory results. When all those forms are marked as "Completed" the patient can be submitted to the medical revisors, who will be able to generate written queries in case of any error or inconsistency, handing back the patient for correction. Every action is performed through the GUI, ensuring full control and guidance. To achieve the most complete data we have adopted two additional approaches:

- Distinction between NULL and Not Available. As example, each field that can be answered with a Boolean logic, should also be able to be answered with "No data available" and NULL, where the latter will indicate a not compiled field and thence an incomplete form. This allows the user to reset safely its input in case of uncertainty and an asynchronous filling of the same form.
- Most fields with structured data include an "Other" option, i.e., a free text option.

The intrinsic validation mechanisms of web forms, e.g., type compliance or value ranges, could not handle these new implementations, so we developed a higher-level protocol, which operates only after the automatic validation assured the low-level compliance. This validator consists in pure code that performs the additional logic checks required, e.g., if a user chose the "Other" option, they must write the answer in the textbox that bears that value. Whenever a form is saved, the validator determines whether it is Complete or not, and the status is represented through the GUI with a colored indicator. Being based on ad hoc code, this validator includes form-dedicated data examination from a medical point of view, achieving a multi-disciplinary validation. This operative design has been already adopted in 3 multicenter studies with a total of 1200 complete and revised patients from 35 unique participating centers, and 2 more are in development.

4. Conclusion

In conclusion, eCRFs proved themselves to be exceptionally efficient in the management of broad and complex clinical data. Our experience proved that a pervasive collaboration between medical researchers and bioengineers is needed to consciously achieve the most complete and meaningful result. In this poster, the authors want to present the operative approach that resulted from this collaboration, highlighting the main points that constitute it, and the advantages and drawbacks associated with them.

References

[1] Shortliffe EH, Cimino JJ. Biomedical informatics: computer applications in health care and biomedicine. Springer; 2014. 1152p.
[2] Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspectives in clinical research. 2014 Oct;5(4):159. doi: 10.4103/2229-3485.140555
[3] McFadden E. Management of data in clinical trials. John Wiley & Sons; 2007 Dec 14. 202p.
[4] Di Meco G, Mora S, Giacobbe DR, Dettori S, Karaiskos I, Bassetti M, Giacomini M. A Wide Database for a Multicenter Study on Pneumocystis jirovecii Pneumonia in Intensive Care Units. Studies in Health Technology and Informatics. 2022 May 1;294:557-8. doi:10.3233/shti220521
[5] Mora S, Giannini B, Di Biagio A, Cenderello G, Nicolini LA, Taramasso L, Dentone C, Bassetti M, Giacomini M. Ten Years of Medical Informatics and Standards Support for Clinical Research in an Infectious Diseases Network. Applied Clinical Informatics. 2023 Jan;14(01):16-27.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230151

Validation Rules as a First Step for Data Quality: Pharmacovigilance Application in Portugal

Bárbara COSTA[a,b], Bruno FIGUEIREDO[b,1] and Vera PINHEIRO[a,b,c]
[a] *CINTESIS, Faculty of Medicine, University of Porto, Portugal*
[b] *MEDCIDS, Faculty of Medicine, University of Porto, Portugal*
[c] *Public Health Unit, Matosinhos Local Health Unit, Matosinhos, Portugal*

Abstract: Tracking and reporting Adverse Drug Reactions (ADRs) is crucial for patient safety. This work aims to improve the data quality of the SIRAI application in Portugal by developing data validation rules and a scoring system for each record and the overall dataset. The goal is to enhance the effectiveness of the SIRAI application in monitoring adverse drug reactions.

Keywords. Data Quality, Pharmacovigilance, Adverse Drug Reaction.

1. Introduction

Adverse drug reactions (ADRs) are potentially preventable responses to medication caused by factors like incorrect administration and communication breakdowns in the healthcare system. [1]. The SIRAI application (Information System for Adverse Reactions and Incidents), at the Matosinhos Local Health Unit (MLHU) is part of the ADR surveillance system in Portugal, but it currently lacks a structured data quality system, essential to ensure accurate reporting of adverse drug reactions [2]. This involves data editing, which detects and corrects errors or inconsistencies in data, usually through data checking validation rules [3]. Thus, the main goal of this work is to contribute to the development of a data quality system for the SIRAI application by establishing checking rules for data validation.

2. Methods

We performed (1) an analysis of the application's guidelines/training materials/interface, the variable list, and the dataset, then (2) we developed checking rules for data validation and (3) we defined a scoring system. It was analyzed a sample of SIRAI dataset, with ADRs reported by MLHU between 2018 and 2022. The checking rules were grouped according to EUROSTAT guidelines [3]: Structural Validation rules, to ensure the technical integrity of the data file, and Content Validation rules, for logical/statistical consistency validation. Finally, we defined a scoring system in accordance with the BANFF methodology [3].

[1] Corresponding Author: Bruno FIGUEIREDO, E-mail: brunofigueiredo@msn.com.

3. Results

From the analysis of the dataset we found that out of the 5250 analyzed entries 35% of non-mandatory variables were missing. Additionally, we identified reporting errors/inconsistencies, which were addressed in the next phase. Based on the dataset structure and issues identified, we established data validation rules - Table 1.

Table 1. Validation rules for the SIRAI Application (some examples)

Rule name	Condition	Error message
Consistency in drug administration dates	Date.end >= Date.beginning	The end of the drug administration must be after the beginning of the drug administration
Consistency in number of ADR-severity	N ADR = N Severity	For N adverse events recorded, there must be exactly the same N records for severity
Completeness of administration route record	VIA_ADMIN AND DESIGNATION != N/A	Drug administration route cannot be missing
Consistency between drug reintroduction and repeated reaction	IF Reintroduction_same_medication="No" THEN Similar_Reaction_Reintroduction="No"	When the suspected drug is not reintroduced, there cannot be a reaction upon reintroduction
Range check for Height	30 cm <= Height <= 250 cm	Height must be between 30 and 250 cm
Range check for Weight	1000 g <= Weight <= 600000 g OR 1 kg <= Weight <= 600 kgs	Weight must be between 1000g and 600000 g (or 1kg and 600 kgs)

ADR - Adverse Drug Reaction; N - Number

We then defined a scoring system. Each record is assigned a status code (PASS, MISS of FAIL) based on the validation rule. Then, an overall record status is derived. Using this method, we can score each recorded ADR and the overall dataset according to the defined validation rules.

4. Discussion and Conclusions

Our goal was to develop a data quality system for the SIRAI application by establishing checking rules for data validation. Data editing systems are essential for the overall quality of any pharmacovigilance system. The defined validation rules and scoring system will be accessible by both managers and reporters of the application, for analyzing data quality issues. In conclusion, implementing a data quality system can improve SIRAI's effectiveness in monitoring ADRs. Future work includes studying the impact of the enhanced data quality system on ADR reporting accuracy and completeness.

References

[1] Coleman JJ, Pontefract SK. Adverse drug reactions. Clin Med (Lond). 2016 Oct;16(5):481-485. doi: 10.7861/clinmedicine.16-5-481. PMID: 27697815; PMCID: PMC6297296.

[2] Ferreira-da-Silva R, Ribeiro-Vaz I, Silva AM, Marques J, Polónia JJ. Looking back on 20 years of work at the Porto Pharmacovigilance Centre Portugal; Cad Saúde Pública. 2021; 37.

[3] EUROSTAT, European Statistical System. Handbook of Methodology for Data Validation 2.0 - revised edition. 2018. In: European Statistical System. [online]. Available from: https://ec.europa.eu/eurostat/cros/system/files/methodology-handbook-on-data-validation-2.0_en.pdf.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230152

A Platform Promoting Inter-Physician Interaction to Support the Management of Adverse Drug Reactions for CLL Patients

Panos BONOTIS[a] and Pantelis NATSIAVAS[a,1]
[a]*Institute of Applied Biosciences, Centre for Research and Development Hellas, Thessaloniki, Greece*
ORCiD ID: Panos Bonotis https://orcid.org/0000-0002-9317-547X
Pantelis Natsiavas https://orcid.org/0000-0002-4061-9815

Abstract. Adverse Drug Reactions (ADRs) cause significant impact for patients' Quality of Life (QoL) and vastly increase costs, especially regarding chronic diseases. To this end, we propose a platform that aims at supporting the management of patients with Chronic Lymphocytic Leukemia (CLL), via an eHealth platform facilitating inter-physician interaction and the provision of treatment consultation by a specialized ADR management team comprised of CLL experts.

Keywords. Chronic Lymphocytic Leukemia, adverse drug reaction, drug-drug interactions

1. Introduction

Chronic Lymphocytic Leukemia (CLL) is the most prevalent adult leukemia and typically occurs in adults with a median age of 72. Currently, treating CLL does not lead to a complete cure, hence a key goal in patient care has been improving patients' quality of life (QoL) [1]. CLL treatment and adverse drug reactions (ADRs) and drug-drug interactions (DDIs) impact the physical and emotional patient status, resulting in anxiety, depression and/or fear of death [2]. They also significantly impact patient therapy adherence and can lead to treatment stop [3], sometimes even without discussion with their physician [4].

2. Platform outline

Along these lines, we propose a platform that aims at supporting the management of CLL ADRs, emphasizing on peer-to-peer (P2P) communication between physicians and a specialized ADR "support" team. The main goals of the platform can be summarized as follows: (a) support regarding the management of ADRs during every-day clinical practice, (b) communication of best practices based on real-world experience and (c) assessment of the impact of the ADRs and CLL treatment as a whole on patient's QoL.

[1] Corresponding Author: Pantelis Natsiavas, eHealth Lab, INAB-CERTH, 6th Km. Charilaou-Thermi Road, Thermi, Thessaloniki, Greece; E-mail: pnatsiavas@certh.gr.

To ensure acceptance, healthcare professionals (HCPs) are engaged in the various stages of the design and the development of such a platform. The proposed technical solution (Figure 1) will consist of three modules, i.e., the *Private Communication Channel (PCC), the Public Forum (PF), and the Analytics Module (AM)*.

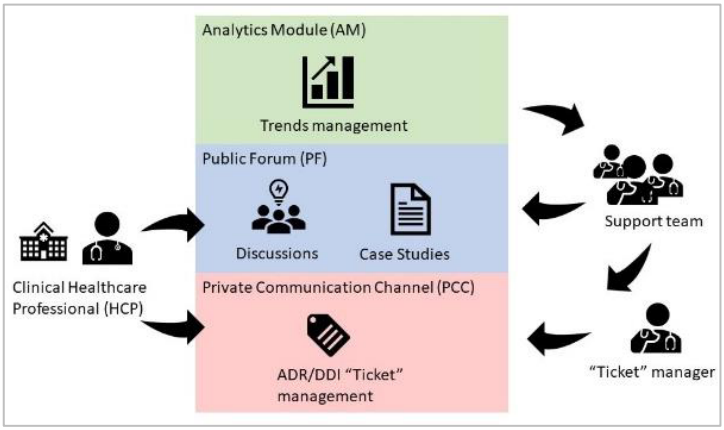


Figure 1. Proposed platform rationale.

The PCC enables direct and private communication between the physicians and the support team (ST). A Ticket Manager (TM) is notified as soon as a support "ticket" is created and he/she coordinates ST's response to the specific ticket. The PF enables the publication of case studies and relevant guidelines by the ST and online discussion functionalities. The AM enables the monitoring of the platform's use and the identification of important ADRs/DDIs and relevant trends.

3. Conclusions

Concluding, we believe that such a digital approach for CLL patient management could be significantly improved by eHealth interventions, emphasizing on ADR management.

References

[1] M. Else *et al.*, "Quality of life in chronic lymphocytic leukemia: 5-Year results from the multicenter randomized LRF CLL4 trial," *Leuk. Lymphoma*, vol. 53, no. 7, pp. 1289–1298, Jul. 2012, doi: 10.3109/10428194.2011.649479.

[2] D. Eek *et al.*,, "Development of a Conceptual Model of Chronic Lymphocytic Leukemia to Better Understand the Patient Experience," *Patient*, vol. 14, no. 1, Jan. 2020, doi: 10.1007/s40271-020-00440-9.

[3] A. Cuneo *et al.*, "Management of adverse events associated with idelalisib treatment in chronic lymphocytic leukemia and follicular lymphoma: A multidisciplinary position paper," *Hematological Oncology*, vol. 37, no. 1. John Wiley and Sons Ltd, pp. 3–14, Feb. 01, 2019, doi: 10.1002/hon.2540.

[4] A. E. Hall *et al.*, "To adhere or not to adhere: Rates and reasons of medication adherence in hematological cancer patients," *Crit. Rev. Oncol. Hematol.*, vol. 97, pp. 247–262, Jan. 2016, doi: 10.1016/j.critrevonc.2015.08.025.

[5] C. Waweru *et al.*, , "Health-related quality of life and economic burden of chronic lymphocytic leukemia in the era of novel targeted agents," *Curr. Med. Res. Opin.*, vol. 36, no. 9, pp. 1–15, 2020, doi: 10.1080/03007995.2020.1784120.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230153

From FAIR4Health Project to 1+MG Initiative: A Spain – Italy Collaboration

Ariam BOAGLIO[a], Carlos Luis PARRA-CALDERÓN[b], Alicia MARTINEZ-GARCÍA[b], María José ESCALONA-CUARESMA[c], Norbert MAGGI[a], Domenico COVIELLO[d], Paolo UVA[d] and Mauro GIACOMINI [a,1]
[a]*DIBRIS, University of Genoa, Genoa, Italy*
[b] *Virgen del Rocio University Hospital - Institute of Biomedicine of Seville, Spain*
[c]*University of Seville, Spain*
[d] *IRCCS Istituto Giannina Gaslini, Genoa, Italy*

Abstract. Results of two major projects funded by the European Union are taken into consideration: Fair4Health regarding the possibility of sharing clinical data in various environments applying FAIR principles and 1+Million Genome for the in-depth study of the human genome in Europe. Specifically, the Gaslini hospital plans to move on both areas joining the Hospital on FHIR initiative matured within the fair4health project and also collaborate with other Italian healthcare facilities through the implementation of a Proof of Concept (PoC) in the 1+MG. The aim of this short paper is to evaluate the applicability of some of the tools of the fair4health project to the Gaslini infrastructure to facilitate its participation in the PoC. One of the aims is also to prove the possibility of reuse the results of well-performed European funded projects to boost routine research in qualified healthcare facilities.

Keywords. FAIR, genome, data sharing, FHIR, translational health informatics.

1. Introduction

In recent years, it has become clear the potential impact and importance of the application of FAIR Principles [1] under the paradigm of open science and of important impact in the health domain. The interest in applying FAIR principles in health care is demonstrated by the development of many research projects funded by the European Union. Among them, one of the most interesting in recent years has been FAIR4Health [2], whose main objective is to make clinical data from different European sites available for to facilitate the European health research community's common use of such data through the application of the FAIR principles. Another very promising line of research in this decade is human genomics. In this area, one of the most relevant recent projects and of a very relevant strategic scale is 1+Million Genome (1+MG) [3], launched in 2018, which was created with the aim of enabling secure access to genomics and corresponding clinical data across Europe to improve research, personalized healthcare, and healthcare policymaking. The Gaslini Pediatric Scientific Hospitalization and Care Institute (IRCCS) plans to move on both areas joining the Hospital On FHIR initiative [4] based on the tools [5] and the HL7 FHIR4FAIR FHIR Implementation Guide [6] matured

[1] Corresponding Author: Mauro Giacomini, E-mail: mauro.giacomini@dibris.unige.it.

within the fair4health project and also to participate in collaboration with other Italian healthcare facilities through the implementation of a Proof of Concept (PoC) in the 1+MG. The aim of this poster is to formulate the applicability of some of the tools of the FAIR4Health project on the Gaslini infrastructure to facilitate its participation in the PoC.

2. Methods

The main outcome of FAIR4Health has been to develop a set of open software tools, among them the ones we want to evaluate useful for our paper are the Data Curation Tool and the Data Privacy Tool, based on the HL7 FHIR standard. The first tool is a standalone desktop application that runs on major operating systems. This application connects to a FHIR endpoint in order to map local existing clinical data into hl7fhir resources. It is possible to select the FHIR profile and to perform mapping operation between the clinical data of the Hospital datasets with the FHIR resources. The dataset and its metadata are curated and validated and then the data privacy tool is used applying de-identification algorithms [7]. For participation in the PoC mentioned above one of the components used by other European groups within 1+MG initiative is Beacon [8], a standard and a service provided by the GPAP (Genome-phenome Analysis Platform) along with its Data Discovery functionality. It allows users to query a controlled access resource and it also allows queries based on the presence of a particular allele at a certain genomic locus [9]. Along with Beacon they used the GA4GH standard Phenopackets [10], a phenotypic data exchange protocol mapping HL7 FHIR to support exchange of phenotypic information.

3. Results

An analysis of the set of information that are necessary for their correct use is underway within a real Italian hospital environment. The authors try to insert these pipelines in the interoperability flow already present in the hospital taking into account all the various levels, physical, semantic and process. Specifically, most of the research documentation is available in the HL7 CDA format and the authors intend to set up elements and criteria for the correct use of this already structured data, tacking account the correct profiling of FHIR resources, both mapping with Phenopackets as the fulfilment of FAIR principles HL7 FHIR4FAIR FHIR IG-based.

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018–160018.
[2] https://www.fair4health.eu
[3] https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes
[4] https://www.hospitalsonfhir.eu/
[5] Sinaci AA, Núñez-Benjumea FJ, Gencturk M, et al. From raw data to fair data: the fairification workflow for health research. Methods Inform Med. 2020;59(S 01):21–32. doi: 10.1055/s-0040-1713684.
[6] http://build.fhir.org/ig/HL7/fhir-for-fair/
[7] https://github.com/fair4health
[8] https://www.ga4gh.org/
[9] Spalding D., Nyrönen T. et al., Secure cross-border data access roadmap (version 23/11/2021)
[10] https://phenopacket-schema.readthedocs.io/en/latest/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230154

Barriers, Priorities and Lessons Learned in Achieving Electronic Health Records Interoperability in Low- and Middle-Income Countries: Workshop Findings

Taiwo ADEDEJI[a,1], Haythem NAKKAS[a], Hamish FRASER [b] and Philip SCOTT [c]
[a] *University of Portsmouth, UK*
[b] *Brown University, UK*
[c] *University of Wales Trinity Saint David, UK*
ORCiD ID: Taiwo Adedeji https://orcid.org/0000-0002-7201-7202

Abstract. Electronic health records (EHR) interoperability is a complex topic that continues to gain traction in the digital health landscape. We facilitated a qualitative workshop consisting of domain experts in EHR implementation and health IT managers. The workshop aimed to identify critical barriers to achieving interoperability, priorities for new EHR implementations and lessons learned from managing existing implementations. The workshop highlighted that data modelling and interoperability standards are vital priorities for maternal and child health data services in low- and middle-income countries (LMICs).

Keywords. Electronic health records, interoperability, data sharing, workshop

1. Introduction

Electronic health records (EHR) interoperability is a complex topic that continues to gain traction in the digital health landscape. Recent studies note that the exchange of captured data via information systems in the health domain can be difficult due to the complex nature of health and social care processes [1,2]. Hence, this contribution reports and reflects on the findings of an expert-led workshop regarding EHR interoperability.

2. Methods

We facilitated a *90-minutes qualitative* workshop at the MedInfo2019 conference. The workshop comprised *EHR implementation findings, lead questions and expert discussions*. The workshop questions are: (1) What are the priorities or drivers for actual EHR implementation? What can we learn from high income countries (HICs) and the few LMICs that are leading in this respect? (2) What are the preconditions for the envisioned EHR implementation, and how can they be met?

[1] Corresponding Author: Taiwo Adedeji, E-mail: taiwo.adedeji@port.ac.uk.

3. Findings

Twenty participants from various institutions attended the session. Below are summary findings from the group discussion: Participants noted many available models with differences in data element specifications, clinical concepts, relationships, and representation formats. For example, one contributor stated that *"there are no centralised data models for most EHR implementations"*. A recurring theme is summarised in the quotation: *"health institutions do not want to share their data models, especially in the US"*. Not only are people reluctant to share their data, but they are reluctant to share their data structures owing to perceived consequences in doing so.

There was a consensus that increased funding for infrastructure is required to assure successful EHR implementation and interoperability with other information systems in LMICs. Drivers for an interoperable EHR differ between LMICs and HICs. For example, LMICs focus mainly on aggregate data from the health information system for disease control, population health monitoring and health policy and planning.

We identified some preconditions for achieving interoperability across EHRs in LMICs: (1) National identifier for patients or citizens. (2) Incentives for EHRs from the government for health institutions. (3) Infrastructure – funding partners, vendors, insurance companies, pension funds (as investors) can support the government (e.g., the Ministry of Health). (4) Clearly defined, evidence-based use cases, such a transmission of laboratory results. These findings are consistent with how the theory of change (ToC) approach could be used to foster interoperability of EHRs, especially in LMICs [3].

4. Conclusion

The workshop highlighted that data modelling and interoperability standards are vital priorities for maternal and child health data services in LMICs. Hence, it is vital to ensure a robust standardisation to collect and share data at all levels. The following summary points were the reflections from the exploratory workshop: (1) Well designed, common data models are vital to realising effective data exchange between digital health systems by adopting the appropriate interoperability standards for MCH. (2) Drivers for an interoperable EHR differ between LMICs and HICs. (3) Adequate infrastructure and accountable funding are essential preconditions for any sustainable EHR implementation, which were not met in the related study [3].

References

[1] Dobrow MJ, Bytautas JP, Tharmalingam S, Hagens S. Interoperable electronic health records and health information exchanges: systematic review. JMIR Medical Informatics. 2019 Jun 6;7(2):e12607.
[2] Johnson KE, Kamineni A, Fuller S, Olmstead D, Wernli KJ. How the provenance of electronic health record data matters for research: a case example using system mapping. EGEMS. 2014;2(1).
[3] Adedeji T, Fraser H, Scott P. Implementing Electronic Health Records in Primary Care Using the Theory of Change: Nigerian Case Study. JMIR medical informatics. 2022 Aug 11;10(8):e33491.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230155

Harvesting the Low Hanging Fruits From the FAIRtree

Esther Thea INAU[1], Atinkut ZELEKE and Dagmar WALTEMATH
Medical Informatics Laboratory, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

Abstract. Extensive workflows have been designed to FAIRify data from various domains. These tend to be cumbersome and overwhelming. This work summarises our own experiences with FAIRification in health data management and provides simple steps that can be implemented to achieve a relatively low but improved level of FAIRness. The steps lead the data steward to register the data in a repository and then annotate it with the metadata recommended by that repository. It further leads the data steward to provide the data in a machine-readable format using an established and accessible language, establish a well-defined framework to describe and structure the (meta)data as well as publish the (meta)data. We hope that following the simple roadmap described in this work helps to demystify the FAIR data principles in the health domain.

Keywords. FAIR, Simple Steps

1. Introduction

Comprehensive and transparent data stewardship involves making data findable, accessible, interoperable and reusable (FAIR) [1]. Canonical workflows have been developed to FAIRify data from specific domains such as geoscience and systems biology [2–4]. This work describes a few simple and helpful steps in the start of a FAIRification journey to achieve a relatively low but improved level of FAIRness in the health domain.

2. Simple Steps in a FAIRification Journey

To improve the findability, we share the dataset via a searchable data repository which will provide a persistent identifier, as well as versioning and tagging with keywords. We also generously enrich the dataset with metadata and register the metadata in preprint servers such as Zenodo which allows for DOI versioning [5]. We make efforts to provide related readme and provenance information alongside the dataset. Registering the data and the associated metadata in searchable data repositories and preprint servers also allows the metadata to remain accessible even after the data is no longer available. We encourage the data owners to decide on the license to choose for the dataset. We then publish the metadata in data repositories such as the Medical Data Models portal to allow

[1] Corresponding Author: Esther Inau, E-mail: inaue@uni-greifswald.de.

downloading and exporting the file in common technical formats and in a standardized manner [6]. We then add contextual knowledge to the dataset in the form of meaningful links to related datasets. Our final inclusion of qualified references increases the interoperability. We then attach rich metadata and standardized vocabularies to the data in a manner that allows to decipher the origin, lineage, usefulness, relevance and how to cite the data in the said context. We also employ domain-specific standards, i.e. (meta)data has the same type, standardized, follows a community accepted template, contains the same type of data organized in a standardized way and uses a common vocabulary in well-established and sustainable file formats.

3. Discussion and Conclusion

Indulging in the FAIRification journey typically leads the data steward to annotate the said data with codes from domain-specific repositories. Through this journey the (meta)data is provided in a machine-readable format using an established and accessible language. This further leads the data steward to establish a well-defined framework to describe and structure the (meta)data in order to facilitate its findability and interoperability. The annotated dataset is then published which increases its reusability the annotated dataset. This journey also leads data stewards to register the (meta)data in the data repositories and preprint servers to enable version control and for researchers to have access to it both now and in the future. This work serves as a basic roadmap that contains small but useful FAIRification steps. The process of FAIRification is far more intensive but for this work but we have chosen to describe only the simpler steps that can be taken to improve data FAIRness. We hope this work helps to demystify the FAIR data principles in the health domain and motivates various stakeholders to start the FAIRification journey by taking the introductory steps described in this work.

References

[1] Suver C, Thorogood A, Doerr M, Wilbanks J, Knoppers B. Bringing Code to Data: Do Not Forget Governance. J Med Internet Res [Internet]. 2020 Jul 28 [cited 2023 Jan 10];22(7):e18087. Available from: http://www.jmir.org/2020/7/e18087/

[2] Wittenburg P, Hardisty A, Le Franc Y, Mozaffari A, Peer L, Skvortsov NA, et al. Canonical Workflows to Make Data FAIR. Data Intell [Internet]. 2022 Apr 1 [cited 2023 Jan 10];4(2):286–305. Available from: https://direct.mit.edu/dint/article/4/2/286/109842/Canonical-Workflows-to-Make-Data-FAIR

[3] Koymans MR, Hinsbergen DJJ, Pastor‐Galán D, Vaes B, Langereis CG. Towards FAIR Paleomagnetic Data Management Through Paleomagnetism.org 2.0. Geochem Geophys Geosystems [Internet]. 2020 Feb [cited 2023 Jan 10];21(2). Available from: https://onlinelibrary.wiley.com/doi/10.1029/2019GC008838

[4] Reiter T, Brooks† PT, Irber† L, Joslin† SEK, Reid† CM, Scott† C, et al. Streamlining data-intensive biology with workflow systems. GigaScience [Internet]. 2021 Jan 13 [cited 2023 Jan 10];10(1):giaa140. Available from: https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giaa140/6092773

[5] European Organization For Nuclear Research, OpenAIRE. Zenodo: Research. Shared. 2013 [cited 2023 Jan 10]; Available from: https://www.zenodo.org/

[6] Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P, et al. Portal of medical data models: information infrastructure for medical research and healthcare. Database [Internet]. 2016 [cited 2023 Jan 10];2016:bav121. Available from: https://academic.oup.com/database/article-lookup/doi/10.1093/database/bav121

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230156

Text Extraction and Standardization System Development for Pathological Records in the Korea Biobank Network

SooJeong KO[a,b] , Sunghyeon PARK[a,b] , SeolWhan OH[a], YunSeon IM[a,b], Surin JUNG[a], BoYeon CHOI[a,b], Jaeyoon Kim[a,b], Wona CHOI [b,1] and InYoung CHOI [b,2]
[a] *Department of Medical Informatics, College of Medicine, The Catholic University of Korea*
[b] *Department of Biomedicine and Health Sciences, The Catholic University of Korea*
ORCiD ID: SooJeong Ko https://orcid.org/0000-0002-6550-9188
Sunghyeon PARK https://orcid.org/0000-0002-2235-4358
SeolWhan OH https://orcid.org/0000-0002-0328-9634
YunSeon IM https://orcid.org/0000-0002-2510-1380
Surin JUNG https://orcid.org/0000-0002-3314-3185
BoYeon CHOI https://orcid.org/0000-0003-1311-2645
Jaeyoon Kim https://orcid.org/0000-0001-8847-9586
Wona CHOI https://orcid.org/ 0000-0003-0269-6374
InYoung CHOI https://orcid.org/ 0000-0002-2860-9411

Abstract. In Korea, the Korea Centers for Disease Control and Prevention operates the Korea BioBank Network (KBN). KBN has pathological records that collected in Korea and it is useful dataset for research. In this study, we established system that time efficient and reduced error by step-by-step data extraction process from KBN pathological records. We tested the extraction process by 769 lung cancer cohorts and 1292 breast cancer cohorts and accuracy is 91%. We expect this system can be used to efficiently process data from multiple institutions, including Korea BioBank Network.

Keywords. NLP, BioBank System,

1. Introduction

Pathological records are essential for many studies such as cancer diagnosis, treatment, and prognosis, but they are usually written in free text format and need to be converted to a standardized and structured format. In Korea, the Korea BioBank Network collects and distributes biobank samples and clinical data [1], but data collection methods vary and manual data entry can lead to errors. Recently, rule-based natural language processing (NLP) systems have been used to extract structured genotype information

[1] Corresponding Author: Wona Choi, E-mail choiwona@gmail.com
[2] Corresponding Author: InYoung Choi, E-mail: iychoi@catholic.ac.kr

from free-text reports [2,3]. To improve data quality, this research finds efficient ways of collecting essential data from pathology records must be found to minimize errors

2. Methods

This study presents a system that uses rule-based extraction to quickly and accurately retrieve data from pathological records texts. The system utilizes Python and its packages to create a user-friendly interface with a step-by-step process that reduces error rates. The process involves creating a sample population, validating and checking for missing data, setting a Large Frame and a Small Frame, saving inserted terms and extracted words as a dictionary, and extracting the final result from the text of the pathological records.

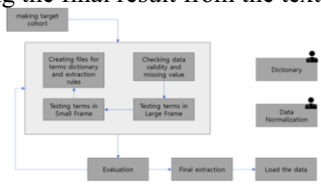


Figure 1. Flow chart of the extraction process

3. Results

This study tested the extraction process on 769 lung cancer and 1292 breast cancer cohorts, achieving an extraction rate of 95% and 91%, respectively. The process was also successful in standardizing the information. Additionally, the time taken for the process was 15 minutes for lung and breast cancer records, as opposed to the previous manual process which took over 1 hour. This suggests a significant reduction in the time required for information extraction.

4. Discussion

This study presents a system for automated data extraction and management from multiple institutions record. The system is designed to allow researchers with knowledge of pathology recording technology and clinical expertise to easily extract and manage data, while allowing the system to check the extraction rate and update extraction conditions for changed terms and forms. Additionally, the system can be installed in medical data management platforms for efficient and effective ETL for quality verification.

References

[1] Ko SJ, Choi W, Kim KH, Lee SJ, Min H, Oh SW, Choi IY. Common Data Model and Database System Development for the Korea Biobank Network. Applied Sciences. 2021 Jan;11(24):11825.
[2] Lee KH, Kim HJ, Kim YJ, Kim JH, Song EY. Extracting structured genotype information from free-text HLA reports using a rule-based approach. Journal of Korean Medical Science. 2020 Mar 30;35(12).
[3] Chen L, et al. Clinical trial cohort selection based on multi-level rule-based natural language processing system. Journal of the American Medical Informatics Association. 2019 Nov;26(11):1218-26.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230157

Refinement of the Target Population for Colorectal Cancer Screening in France, Inventory as a Prelude to the Use of Medico-Administrative Database

Akoi KOÏVOGUI [a,1], Robert BENAMOUZIG [b] and Catherine DUCLOS [a]
[a] *Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des connaissances en e-Santé, LIMICS, Sorbonne Université, UMLS 1142-INSERM, 93000, Bobigny, France*
[b] *AP-HP, Avicenne, Service d'Hépato Gastro Entérologie, 93000, Bobigny, France*

Abstract. The study describes the level of improvement in the risk of misclassification that would be achieved by refining the campaign target population using a query in the French medico-administrative database (SNDS). The SNDS's use requires other new strategies that can minimize the number of people wrongly included in the campaigns, because its accuracy is less than 100%.

Keywords. Medico-administrative database, Colorectal cancer screening

1. Introduction

The French colorectal cancer (*CRC*) screening program (*CRCSP*) targets people aged 50 to 74. The medical exclusion (*MEx*) rate and the participation rate in 2020-2021 (respectively: 13.2% 35.7%; *www.santepubliquefrance.fr*) are estimated in an eligible population (*EliP*) and underestimated. To obtain *EliP*, a few people (i.e., people at risk of *CRC*) are excluded from target population (*TaP*). Unlike some European programs [1], in France, the *MEx* is assessed after invitations (Response from patients/GPs/physicians), generating a detrimental bias. With a view to interconnecting the health insurance medico-administrative database (*SNDS*) and the *CRCSP* database (*CRC-DB*), we aim to describe the risk of misclassification (Eligible/Not eligible) induced by *MEx* strategies and the level of risk's improvement that, would be achieved by refining the *TaP* upstream of invitations using *SNDS*.

2. Methodology

The CRCSP's major *MEx* strategies are:

[1] Corresponding Auhtor: Dr KOÏVOGUI Akoï, CRCDC-IDF, Seine-Saint-Denis site, 41 avenue de Verdun, 93146 Bondy, France. E-mail: akoi.koivogui@inserm.fr Tel.: +33155 89 10 10 Fax+33148 02 06 80

i) *MEx* made exclusively upstream of the invitations by *CRCSP'*s management structures (*CRCSP-Manager*) through the disparate use of: *CRC-DB*, Cancer registries, Hospital discharge databases and Primary health insurance databases. ***ii)*** *MEx* made following correspondence between *CRCSP-Manager* and GPs/Physician through the *Computerized Medical Record* (*CMR*) of the physician/healthcare facility. ***iii)*** *MEx* carried out by pharmacist by answering the questions of eligibility when collecting a test kit at the pharmacy. ***iv)*** The self-exclusion carried out by people either by making a postal response, or by completing an online eligibility form. ***v)*** *MEx* made through the *SNDS* (In perspective). Each strategy's risk of misclassification was analyzed in terms of the proportion of people potentially eligible for screening and included appropriately (*TP*: True positive) or wrongly excluded (*FN*: False negative) and, in terms of the proportion of people potentially ineligible for screening and wrongly included (FP: False positive) or appropriately excluded (*TN*: True negative). These *FP*, *FN, TN*, and *TP* were simulated in five scenarios, based on the five *MEx* strategies.

3. Results

To obtain a test kit, people must: ***Scenario-1 (Doctor consultation)***: eligibility by strategies ***(i)*** and ***(ii)*** with zero risk of misclassification (*TP*=100%, *FP*=0%, *FN*=0%, *TN*=100%)**.** ***Scenario-2(Kit collect in pharmacy)***: There is a non-zero risk that eligibility (Strategy ***iii)*** may be partially proved (*TP*<100%, *FP*>0%, *FN*=0%, *TN*=100%) due to pharmacist 's lack of access to the *CMR* or *CRC-DB*. ***Scenario-3(Kit request, no consultation)***: The kit is received at home following a website request, eligibility is established (Strategy ***iv)*** with a non-zero risk (*TP*<100%, *FP*>0%, *FN*>0%, *TN*<100%). ***Scenario-4 (SNDS perspective)*** risks are like those of ***Scenario-3,*** because SNDS accuracy <100%. The risk that few people will be wrongly included or not included, in the participation rate's estimation is >0 in scenarios 2, 3 and 4 and =0 in 1.

4. Discussion and Conclusions

The *SNDS*'s sensitivity (≤87.7%) and positive predictive value(≤58.9%) are less than 100 [2]. This inaccuracy feed the reflection with an essential question: To what extent refining upstream invitations, by a *SNDS* query will improve the exclusion rate?

To making *MEx* using the SNDS, other new strategies are needed to minimize the number of people not wrongly included in the campaigns. The definition of these new strategies will have to go through a *SNDS's* evaluation using *CRC-DB* as a benchmark.

References

[1] Bulliard JL, Garcia M, Blom J, et al. Sorting out measures and definitions of screening participation to improve comparability: the example of colorectal cancer. Eur J Cancer. 2014;50(2):434-46.

[2] Quantin C, Benzenine E, Hagi M, et al. [Evaluation of medical information systems as a mean of identification of incident cases of colorectal cancer]. Sante publique 2014;26(1):55-63.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230158

Transfer Learning for Early Prediction of Adverse Drug Reactions: Docetaxel and Alopecia in Breast Cancer as a Case Study

Stella DIMITSAKI[a,b,1], Pantelis NATSIAVAS[b] and Marie-Christine JAULENT[a]
[a] *Sorbonne Université, INSERM, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la eSanté, LIMICS, F-75006 Paris, France*
[b] *Institute of Applied Biosciences, Centre for Research and Development Hellas, Thessaloniki, Greece*
ORCiD ID: Stella Dimitsaki https://orcid.org/0000-0002-6356-9531
Pantelis Natsiavas https://orcid.org/0000-0002-4061-9815
Marie-Christine Jaulent https://orcid.org/0000-0003-4445-7494

Abstract. Transfer Learning (TL) is an approach which has not yet been widely investigated in healthcare, mostly applied in image data. This study outlines a TL pipeline leveraging Individual Case Safety reports (ICSRs) and Electronic Health Records (EHR), applied for the early detection Adverse Drug Reactions (ADR), evaluated using of alopecia and docetaxel on breast cancer patients as use case.

Keywords. Breast cancer, docetaxel, adverse drug reaction, alopecia, transfer learning

1. Introduction

Transfer Learning (TL) is a Machine Learning (ML) approach where a trained model for a specific task/dataset is retrained for another task/dataset. In healthcare, TL is currently mainly used for prediction models built upon image data [1]. Recently, TL is introduced also for Adverse Drug Reaction (ADR) prediction [2]. Our study aims to develop a TL predictive approach upon Individual Case Safety reports (ICSRs) and Electronic Health Records (EHR), applied for the early detection of ADRs signals, typically elaborated through disproportionality analysis applied in Pharmacovigilance (PV) pipelines. The presented approach is evaluated using alopecia, an ADR observed following chemotherapy with docetaxel in BC patients. An early detection of this ADR could help BC patients to confront anxiety and distress via proper psychological preparation and other potential mitigation measures (e.g., scalp cooling) [3].

[1] Corresponding Author: Stella Dimitsaki, E-mail: sdimitsaki@certh.gr.

2. Methods

ICSR data from the FDA Adverse Event Reporting System (FAERS) will be used for the initial training step and the retrained step will be applied upon EHR data. Finally, the trained TF model will be evaluated in terms of predicting the adverse effect of alopecia on BC patients or the identification of the respective PV signal using the combination of ICSR and EHR data, compared to the use of tradition PV methods.

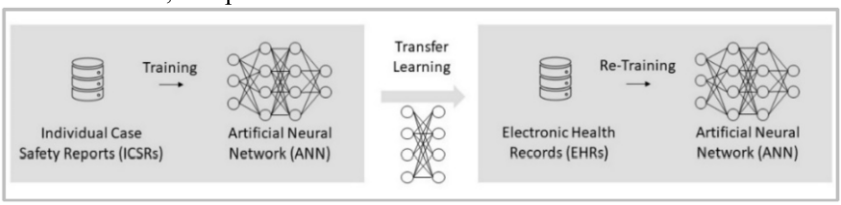


Figure 1. Methodology rationale.

ICSR data are preprocessed, i.e. "cleaned" to facilitate further processing. Relevant information provided in ICSR data is taken into account, e.g. age and suspected drugs. Indicatively, Figure 2 depicts the number of concomitant suspected active substances in the selected ICSRs. Furthermore, ATC codes and the relevant classification of active substances is used to identify potential associations for drug classes.

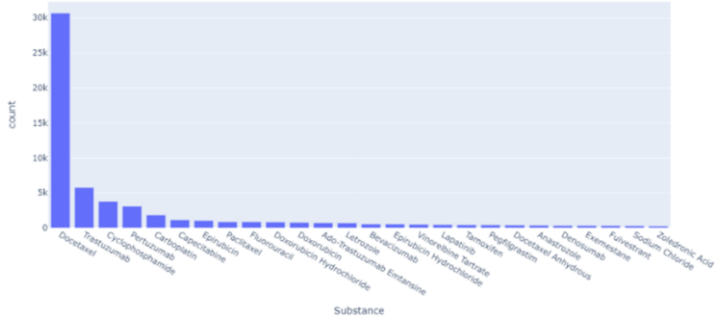


Figure 2. Top 26 concomitant/suspected active substances in the Docetaxel's ICSRs in FAERS

3. Conclusions

Based also on the outcome of other works combining both ICSRs and EHR data [4], we argue that such an "intelligent" approach could provide speeding the detection of ADRs.

References

[1] Ebbehoj A, et al. Transfer learning for non-image data in clinical research: a scoping review. PLOS Digital Health. 2022 Feb 17;1(2):e0000014. doi: 10.1371/journal.pdig.0000014.

[2] Ekpenyong ME, et al. A transfer learning approach to drug resistance classification in mixed HIV dataset. Informatics in Medicine Unlocked. 2021 Jan 1;24:100568. doi: 10.1016/J.IMU.2021.100568.

[3] Martín M, et al. Persistent major alopecia following adjuvant docetaxel for breast cancer: incidence, characteristics, and prevention with scalp cooling. Breast Cancer Research and Treatment. 2018 Oct;171:627-34., doi: 10.1007/s10549-018-4855-2.

[4] Yu Y, et al. ADEpedia-on-OHDSI: A next generation pharmacovigilance signal detection platform using the OHDSI common data model. Journal of biomedical informatics. 2019 Mar 1;91:103119.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230159

An Audit of Dietitians' Documentation – Comparing the Level of Agreement Between the Audit Instruments Diet-NCP-Audit and NCP-QUEST

Anja LINDBERG[a], Maria NYBERG[b], Ylva ORREVALL[c] and Evelina LILJEBERG[c,d,1]
[a] *Nacka Rehab Centre, Stockholm, Sweden*
[b] *Primary Health Care Dietitians, County Västmanland, Sweden*
[c] *Medical Unit Clinical Nutrition, Women's Health and Allied Health Professionals Theme, Karolinska University Hospital, Stockholm, Sweden*
[d] *Department of Food Studies, Nutrition and Dietetics, Uppsala University, Sweden*
ORCiD ID: Evelina Liljeberg https://orcid.org/0000-0002-2605-7457

Abstract. To explore the inter-rater reliability of the Swedish translation of NCP-QUEST in a Swedish context and investigate the level of agreement between Diet-NCP-Audit and NCP-QUEST in assessment of documentation quality. A retrospective audit was conducted of 40 electronic patient records written by dietitians at one University Hospital in Sweden. NCP-QUEST showed good inter-rater reliability for the quality category (ICC = 0.85) and excellent inter-rater reliability for total score (ICC = 0.97).

Keywords. Documentation Quality, Medical Record, Dietitian, Nutrition Care Process

1. Introduction

The dietetic profession is young and how the nutrition care provided by dietitians should be documented in patient's health records is under constant evolvement. During the last decade, the Diet-NCP-Audit [1] has been the only validated audit tool available to evaluate the quality of dietitians' documentation. However, in 2021, a new validated audit tool was published: NCP Quality Evaluation and Standardization Tool (NCP-QUEST) [2]. NCP-QUEST is an update of Diet-NCP-Audit and both instruments are designed to evaluate if the documentation covers the four steps of the Nutrition Care Process (NCP) which is a profession specific working process with an accompanying terminology [2, 3]. NCP-QUEST comprises some NCP terminology updates and was recently translated into Swedish, but its inter-rater reliability and agreement with Diet-NCP-Audit is yet to be explored. To explore the inter-rater reliability of the Swedish translation of NCP-QUEST in a Swedish context and investigate the level of agreement between Diet-NCP-Audit and NCP-QUEST in assessment of documentation quality.

[1] Corresponding Author: Evelina Liljeberg, E-mail: evelina.liljeberg@ikv.uu.se.

2. Method

A retrospective audit was conducted of 40 electronic patient records written by dietitians at one University Hospital in Sweden. All patient records were audited with both Diet-NCP-Audit [1] and NCP-QUEST [2] and given a score of the quality of documentation for each instrument. Depending on the score given, the patient records were placed into a quality level (A=high, B=medium, C=low). Subsequently, a Cohens Weighed Kappa analysis was performed to assess the level of agreement between the instruments. A value between 0-0.2 indicate no agreement, 0.21-0.39 minimal, 0.40-0.59 weak, 0.6-0.79 moderate, 0.8-0.9 strong and >0.9 almost perfect agreement [4]. Ten patient records were selected for the assessment of inter-rater reliability between two coders which was made by an intraclass correlation coefficient (ICC) analysis. An ICC value between 0.75-0.9 indicate good reliability and > 0.9 indicate excellent reliability [5].

3. Results, Discussion and Conclusions

NCP-QUEST showed good inter-rater reliability for the quality category (ICC = 0.85) and excellent inter-rater reliability for total score (ICC = 0.97). The audit with Diet-NCP-Audit showed that 70% of the patient records had high quality and 7.5% had low quality (Table 1). When the same dietetic records were audited with NCP-QUEST, only 17.5% had high quality, whereas 47,5% had low quality (Table 1). The two least documented items according to NCP-QUEST concerned if standardized NCP terms were used. Nutrition goal and prescription were the two least documented items according to Diet-NCP-Audit. The level of agreement between the instruments was weak (K value = 0.16).

Table 1. Proportion of diabetic records placed in the three different quality categories by using the audit instruments NCP-QUEST and Diet-NCP-Audit.

	NCP-QUEST		Diet-NCP-Audit	
Quality level	*n*	%	*n*	%
A (High)	7	17,5	28	70
B (Medium)	14	35	9	22,5
C (Low)	19	47,5	3	7,5
Total	40	100	40	100

The Swedish translation of NCP-QUEST is a reliable audit tool and had a weak agreement with Diet-NCP-Audit. This indicates that the criteria for what is considered high documentation quality differ between the two instruments. Even if higher documentation quality is achieved with both instruments when documenting according to NCP, maximum scoring with NCP-QUEST requires a strict adherence to NCP steps and specific terminology.

References

[1] Lövestam E, et al. Evaluation of a Nutrition Care Process-based audit instrument, the Diet-NCP-Audit, for documentation of dietetic care in medical records. Scand J Caring Sci. 2014;28(3):390-397.
[2] Lewis S L, et al. Nutrition Care Process Quality Evaluation and Standardization Tool: The Next Frontier in Quality Evaluation of Documentation. J Acad Nutr Diet. 2021;122(3): 650-660.
[3] Swan WI, et al. Nutrition Care Process and Model Update: Toward Realizing People-Centered Care and Outcomes Management. J Acad Nutr Diet. 2017;117(12):2003-2014.
[4] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med. 2012;22(3):276–282
[5] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–163.

This page intentionally left blank

Section 3

Citizen Health Informatics, Societal Aspects and Education

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230161

What Do Autistic People Discuss on Twitter? An Approach Using BERTopic Modelling

Elia GABARRON[a,b 1] and Enrique DORRONZORO[c], Daniel REICHENPFADER[d] and Kerstin DENECKE[d]
[a] *Department of Education, ICT and Learning, Østfold University College, Halden, Norway*
[b] *Norwegian Centre for E-health Research, Tromsø, Norway*
[c] *Universidad de Sevilla, Seville, Spain*
[d] *Bern University of Applied Sciences, Institute for Medical Informatics I4MI, Biel/Bienne, Switzerland*

Abstract. Social media provide easy ways to autistic individuals to communicate and to make their voices heard. The objective of this paper is to identify the main themes that are being discussed by autistic people on Twitter. We collected a sample of tweets containing the hashtag #ActuallyAutistic during the period 10/02/2022 and 14/09/2022. To identify the most discussed topics, BERTopic modelling was applied. We manually grouped the detected topics into 6 major themes using inductive content analysis: 1) General aspects of autism and experiences of autistic individuals; 2) Autism awareness, pride and funding; 3) Interventions, mostly related to Applied Behavior Analysis; 4) Reactions and expressions; 5) Everyday life as an autistic (lifelong condition, work, housing…); and 6) Symbols and characteristics. The majority of tweets were presenting general aspects and experiences as autistic individuals; raising awareness; and about their dissatisfaction with some interventions. The identification of autistic individuals' main discussion themes could help to develop meaningful public health agendas and research involving and addressed to autistic individuals.

Keywords. Autism Spectrum Disorder, Social Media, Twitter, Topic Modelling, BERT

1. Introduction

Autism Spectrum Disorder is characterized for restricted interests, repetitive patterns of behaviors, and impairments in both, socialization and communication [1]. Despite these impairments, many autistic individuals use social media to socialize [2-3], and to make their voices heard [2-4]. Social media are perceived by some autistic people as easier media for communicating and being social than in-person interactions [4].

Not much is known about what autistic individuals discuss on social media. Research identifying and analyzing discussions related to autism on social media exists

[1] Corresponding Author: Elia Gabarron, Department of Education, ICT and Learning, Østfold University College, B R A Veien 4, 1757 Halden (Norway), E-mail: elia.gabarron@hiof.no

[5-8]. However, the analyzed posts by these studies were non-specific and could have been posted by autistic people, or by other individuals interested in that topic. As far as we know, only few studies have specifically focused on messages posted by autistic individuals [9-11]. These studies have found that autistic social media users utilize these channels to make meaning of their own experiences, such as rejecting medicalized narratives or flipping negative narratives into positive stories [9,10]; and to post about autism characteristics and societal barriers [10].

The use of automatic natural language processing methods could help to identify major themes posted by autistic individuals, and therefore increase our understanding of autistic priorities and concerns. The objective of this paper is to identify the main themes that are being discussed by autistic individuals on Twitter by using BERTopic modelling.

2. Method

2.1. Tweets identification and data extraction

Tweets are published within the public domain, and are frequently used in research [5-11]. Twitter provides Application Programing Interface (API), enabling programmatic extraction of the data. Using this API, we collected a sample of tweets including the hashtag #ActuallyAutistic, among others, during the period 10/02/2022 and 14/09/2022. The limitations of the API do only allow to download 900 tweets every 15 minutes. To download a representative quantity of tweets, we developed a script to run a query several times per day. Together with the text of the tweet we collected detailed user metadata (such as user name, profile image, posting time, and follower number). For the purpose of this study, we filtered the collection of tweets to only include those with the hashtag #ActuallyAutistic. This hashtag was chosen for being the most representative one used by social media users as a means of identifying and revealing themselves as autistic.

2.2. Topic modelling and content analysis

We used the BERTopic modelling technique to identify the main discussed themes [12]. Minor data pre-processing included removal of user handles (starting with "@") and hyperlinks, removal of the hashtag "#ActuallyAutistic", and of common English stop words like "the", "a", or "and". Other hashtags were preserved (by removing the "#") and emojis were decoded to corresponding text representations. Furthermore, contractions where expanded ("I'm" -> "I am"). Hence, tweets that only contained user handles and/or hyperlinks vanished and were not considered for analysis. Applying BERTopic modelling comprises following steps: 1) tweets were converted to vector representations; 2) the dimensionality of these vectors was reduced using the UMAP approach [13]; 3) the data was clustered and bag-of-words representations were generated for each cluster; and 4) the topics were generated using a class-based TF-IDF model. As a language embedding model we used the default sentence-transformer model "all-MiniLM-L6-v2" [14]. The code was run using default parameters, except defined as follows: BERTopic (nr_topics=auto, min_topic_size=300), UMAP (n_neigbours=200).

Inductive content analysis was used to group the major categories identified by the BERTopic modelling technique, which included examples of representative tweets, into main themes. Any disagreements were resolved by discussion.

The treatment of personal information for this study was approved by the data protection officer at the University Hospital of North Norway (Ref:02275).

3. Results

A total of 98601 tweets and retweets including the hashtag #ActuallyAutistic were extracted. Authors of these tweets followed an average of 1570,3 other users, and had a mean of 2440,7 followers. After minor pre-processing, a total amount of 97751 tweets was fed into the BERTopic model. The model auto-generated 35 topics related to the Twitter debate including the hashtag #ActuallyAutistic. The number of topics was further auto-reduced by BERTopic to 20. We manually grouped these 20 topics into six major themes using inductive content analysis. Identified major themes were: T1-General aspects and experiences of autistic individuals; T2-Autism awareness, pride and funding; T3-Interventions, mostly related to Applied Behavior Analysis (ABA); T4-Reactions, expressions; T5-Everyday life as an autistic (lifelong condition, work, housing…); and T6-Symbols, characteristics. The remaining tweets were automatically classified by the BERTopic algorithm into two topics including non-English tweets as well as into an "outlier" class, for which topic assignment could not be done. Included topics in these themes, and number of tweets that were classified into these topics are presented in Table 1.

Table 1. Major themes discussed on tweets including the hashtag #ActuallyAutistic

Theme	Sets of relevant words in each theme [BERTtopic grouped sub-topics]	Tweets
T1: General aspects	['autistic', 'autism', 'people', 'like', 'just', 'know', 'neurodivergent', 'adhd', 'autismacceptancemonth' 'autismacceptance']	44214 (45.2%)
T2: Autism awareness, pride, funding	['april', '*oncomingfist*', '*thumbsup*', 'single', '*smilingfacewithsmilingeyes*', '*blueheart*', 'fact', 'autismawarenessmonth', 'year', 'understand'], ['thread', 'april', 'month', 'pure', 'folx', 'know', 'hell', 'probably', 'aware', 'awareness'], ['donate', 'money', 'donating', 'help', 'donations', 'mutualaidrequest', 'autismacceptancemonth', 'gofundme', 'fundraiser', 'support'], ['day', 'choose', 'express', 'pride', 'autisticprideday', 'autistics', 'today', 'autistic', 'aye', 'june'], ['taxi', 'thai', 'nftart', 'nfts', 'autismpride', 'fake', 'nft', 'nftproject', 'nftcommunity', 'art']	3344 (3.4%)
T3: Interventions, mostly ABA	['lgbtq', 'aba', 'applied', 'analysis', 'lovaas', 'behavioral', 'ivar', 'ole', 'fates', 'intertwined'], ['symbols', 'despised', 'refuse', 'eradication', 'dedicated', 'eugenics', 'fund', 'torture', 'company', 'abuse'], ['month', 'awarenessautism', '*foldedhandslightskintone*', 'saynotoautismspeaks', 'clear', 'word', 'sure', 'thank', 'rainbow', 'infinity'], ['cured', 'cure', 'autism', 'disease', 'accommodation', 'abundantly', 'eradicated', 'eugenics', 'acceptance', 'form']	3309 (3.4%)
T4: Reactions, expressions	['description', 'wow', 'perfect', 'problem', 'come', 'answer', 'induction', 'functiooon', 'whomever', 'compelling'], ['funny', 'laugh', 'reasons', 'situation', 'mean', 'actually', 'does', 'person', 'giggle', 'dying'], ['sometimesoften', 'embarrass', 'solid', 'crucial', 'shared', 'details', 'conversation', 'small', 'communication', 'believe']	3302 (3.4%)

T5: Everyday life as an autistic	['lifetime', 'masking', 'trying', 'unmasked', 'standards', 'exhausted', 'holding', 'spent', 'appease', 'concrete'], ['leave', 'functioning', 'checklist', 'lists', 'require', 'jobs', 'house', 'saw', 'low', 'deserve'], ['*hollowredcircle*', '*crossmark*', 'rent', 'help', '164', 'urgent', 'cut', 'fools', 'badly', 'paying']	2081 (2.1%)
T6: Symbols, characteristics	['boys', 'soon', 'created', 'common', 'coming', 'okay', 'thought', 'true', 'puzzle', 'symbol'], ['symbol', 'infinity', 'puzzling', 'outdated', 'gold', 'fixed', 'condition', 'general', 'meaning', 'offensive']	1750 (1.8%)
Non-English tweets	['ich', 'autismus', 'je', 'und', 'et', 'ist', 'pas', 'le', 'die', 'zu'], ['soyautista', 'que', 'el', 'autismo', 'en', 'por', 'autista', 'la', 'autistas', 'para']	1254 (1.3%)
Outliers	Topics that could not be assigned to one of the above-mentioned topics	38497 (39.4%)
Total		**97751**

* *Emojis are presented in italics*

4. Discussion

Many individuals display their conditions on social media, including autistic individuals. The topic analysis approach we used in this paper provides a tool to learn more about aspects of relevance for autistic social media users. Studying topics discussed by autistic people over time can provide details about recent and current pains and needs of those people. In this way, more focused public health agendas and research can be developed.

In our study we have identified that autistic people on Twitter mostly discuss about general aspects and their experiences as autistic individuals, and about raising awareness. Our findings are in agreement with what has been identified in previous research [9-10]. However, dissatisfaction with some current interventions, specifically the use of ABA, was also identified as one of the most main discussed themes in our database (i.e., 'refuse', 'eradication', 'torture', or 'abuse' were some of the most commonly used keywords linked to ABA interventions). Opposition to the use of ABA in education of autistic people has previously been reported as being part of the neurodiversity movement [15-16]. The community disapproval of this type of interventions should be considered by authorities and researchers when developing future intervention programmes.

The literature shows that the neurodiversity movement also claims increased efforts to develop and validate tools to measure autistic prioritized outcomes [16]. The use of automatic natural language processing methods, and specifically BERTopic could help to identify some of these priorities quickly and on time, and by directly listening to users. Both BERTopic and non-negative matrix factorization have been found to be superior techniques for analyzing Twitter data, in comparison to latent Dirichlet allocation and Top2Vec approaches [17]. Advantages of BERTopic include no to little data pre-processing effort, automatic topic reduction and high stability across domains [17]. Moreover, instead of calculating topic distribution, BERTopic assigns each tweet to exactly one topic. However, a potential high number of auto-generated topics and outliers as well as the absence of objective evaluation metrics have also been reported [17].

Our study presents several limitations. We have only used one hashtag as a way to identify messages posted by autistic individuals on Twitter. Findings from this study cannot be generalized to all autistic individuals, neither to all autistic social media users. There are also some limitations to our chosen approach. We decided not to filter non-English tweets from the input data but used an English resulting in the generation of two

"phantom-topics" containing these non-English tweets. As we did not prevent stochastic behaviour when creating the results described in this paper, we are not able to reproduce the exact same topic representations. Several runs of the modelling pipeline however showed similar results across all runs.

5. Conclusions

BERTopic modelling can be a valid approach to identify main themes that are discussed by autistic individuals on Twitter. Autistic people on Twitter seems to mostly discuss about general aspects and experiences as autistic individuals, raising awareness, and about their dissatisfaction with ABA interventions. Future research could further investigate how topic modelling approaches could help to identify autistic individuals' priorities and concerns on time. Since social media has no borders, but public health interventions are developed in specific areas, future research could also investigate how to transfer this knowledge to develop meaningful public health agendas and research that involve and are addressed to autistic individuals in different regions.

References

[1] American Psychiatric Association 2013. Diagnostic and statistical manual of mental disorders. 5th ed.
[2] Ward DM, Dill-Shackleford KE, Mazurek MO. Social Media Use and Happiness in Adults with Autism Spectrum Disorder. Cyberpsychol Behav Soc Netw 2018;21(3):205-209.
[3] Jones S. TikTok is teaching the world about autism – but is it empowering autistic people or pigeonholing them? Theconversation.com 2022. Available at https://theconversation.com/tiktok-is-teaching-the-world-about-autism-but-is-it-empowering-autistic-people-or-pigeonholing-them-192093
[4] Wayman L. Social media offers benefits for autistic community. Autism speaks 2021. Available at: https://www.autismspeaks.org/science-news/social-media-offers-benefits-autistic-community
[5] Shakes P, Cashin A. An analysis of Twitter discourse regarding identifying language for people on the autism spectrum. Issues Ment Health Nurs 2020;51(3):221-228.
[6] Corti L, Zanetti M, et al. Social media analysis of Twitter tweets related to ASD in 2019-2020, with particular attention to COVID-19: topic modelling and sentiment analysis. J Big Data 2022;9:113.
[7] Gabarron E, Dechsling A, Skafle I, et al. Discussions of Asperger Syndrome on Social Media: Content and Sentiment Analysis on Twitter. JMIR Form Res 2022;6(3):e32752.
[8] Skafle I, Gabarron E, Dechsling A, et al. Online Attitudes and Information-Seeking Behaviour on Autism, Asperger Syndrome and Greta Thunberg. Int J Environ Res Public Health 2021;18(9):4981.
[9] Egner J. #ActuallyAutistic: using Twitter to Construct Individual and Collective Identity Narratives. Stud Soc Justice 2022;16(2):349-369.
[10] Zolyomi A, Jones R, Kaftan T. #ActuallyAutistic sense-making on Twitter. Conference on Computers and Accessibility (ASSETS '20), October 26–28,.2020, Virtual Event, Greece.
[11] Hswen Y, et al. Using Twitter to detect psychological characteristics of self-identified persons with Autism Spectrum Disorder: A feasibility study. JMIR Mhealth Uhealth 2019;7(2):e12264.
[12] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv 2002:2203.05794). Available at: https://doi.org/10.48550/arXiv.2203.05794
[13] McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 2020:1802.03426. Available at: https://doi.org/10.48550/arXiv.1802.03426
[14] Reimers N, Gurevych I. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Available at: https://arxiv.org/abs/2004.09813
[15] Løkke JA, Orm S, Dechsling A. Misconceptions about applied behavior analysis in the neurodiversity movement. Norsk Tidsskrift for Atferdsanalyse 2019;46:55-63.
[16] Leadbitter K, Buckle KL, Ellis C, et al. Autistic self-advocacy and the neurodiversity movement: Implications for autism early intervention research and practice. Front Psychol 2021;12:635690.
[17] Egger R, Yu J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Front Sociol 2022;7:886498.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230162

Scale Up Multilingualism in Health Emergency Learning: Developing an Automated Transcription and Translation Tool

Heini UTUNEN[a,1], Thomas STAUBITZ[b], Richelle GEORGE[a], Yu (Ursula) ZHAO[c], Sebastian SERTH[b] and Anna TOKAR[a]
[a] *Health Emergencies Programme, World Health Organization, Geneva, Switzerland*
[b] *Hasso Plattner Institute, University of Potsdam, Potsdam, Germany*
[c] *Digital health and innovation, World Health Organization, Geneva, Switzerland*

Abstract. World Health Organization's (WHO) emergency learning platform OpenWHO provided by Hasso Plattner Institut (HPI) delivered online learning in real-time and in multiple languages during the COVID-19 pandemic. The challenge was to move from manual transcription and translation to automated to increase the speed and quantity of materials and languages available. TransPipe tool was introduced to facilitate this task. We describe the TransPipe development, analyze its functioning and report key results achieved. TransPipe successfully connects existing services and provides a suitable workflow to create and maintain video subtitles in different languages. By the end of 2022, the tool transcribed nearly 4,700 minutes of video content and translated 1,050,700 characters of video subtitles. Automated transcription and translation have enormous potential as a public health learning tool, allowing the near-simultaneous availability of video subtitles on OpenWHO in many languages, thus improving the usability of the learning materials in multiple languages for wider audiences.

Keywords. digital health, artificial intelligence (AI), online learning, health emergencies.

1. Introduction

OpenWHO is the WHO's open source learning platform in health emergencies, providing unlimited access to health professionals and anybody in need of epidemic and pandemic learning. It is a platform provided and adjusted for the purpose by HPI. The Covid-19 pandemic exponentially increased registered users from 160 000 enrolments in January 2020 to more than 7.5 million currently. The platform provides courses on pandemic and epidemic topics in self-paced modalities with videos, quizzes and reference materials.

During the COVID-19 pandemic, OpenWHO needed to deliver lifesaving training materials to frontline health workers in a timely manner and in their own language.

[1] Corresponding author: Heini Utunen, WHO Health Emergencies Programme, World Health Organization. Avenue Appia 20, Geneva, Switzerland; Geneva, Switzerland; E-mail: utunenh@who.int.

However, quick and reliable translation can be difficult and expensive, often requiring manual and time-intensive work [1]. First, there was a lack of automated systems in place, causing repetitive transcription and translation. Second, there was limited translation capacity due to competing tasks. Moreover, results could be undermined by the complexity of the materials [2]. Even when automated transcription and translation services were utilized, the processes of getting videos to these services, retrieving the results, providing quality assurance (QA), and adding the transcripts and translations as subtitles to the videos was a completely manual process. Several steps in this process were highly error-prone. For example, the subtitle file format in use must be machine readable for display with the video. Particularly, manual editing in word processors or sending the files back and forth per mail often resulted in broken subtitle files or wrong character encodings, which then caused errors with the display of the affected videos. Incremental innovation was therefore urgently needed to tackle the above challenges and to mitigate potential risks of failure during the emergency scaling up. Between April 2020 and September 2021 OpenWHO and Hasso Plattner Institut (HPI) created, tested, and finally established TransPipe, a user interface that facilitates the automated transcription and translation solution with possibility of human interference for quality control. TransPipe itself is not an artificial intelligence tool. It, basically, connects the video provider, the learning platform and several transcription and translation tools, which use artificial intelligence to provide their services. The name TransPipe is short for "Transcription and Translation Pipeline". Its purpose is to fill the bits in between the existing tools and to provide a smooth and streamlined process that eliminates the pitfalls of the previous manual workflow. On top of connecting the existing tools, TransPipe adds a quality assurance workflow that allows human editors, translators, and quality assurance managers to be assigned to a certain course, video, or language. Finally, it visualizes the current state of the process of a certain course, video, or language in a course (see Figure 1).

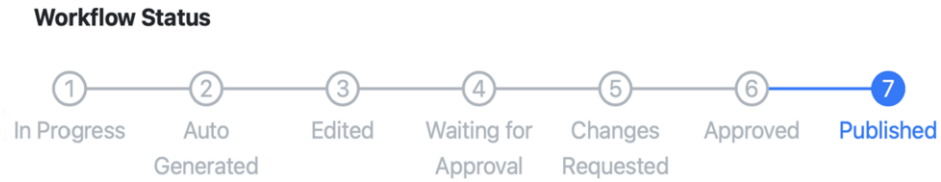


Figure 1. Visualization of workflow status

This article aims to offer an insight into the TransPipe [3] development from design, test and deployment, followed by a brief summary of the tool's performance since its deployment. The article also discusses the limitations and potentials for upcoming iterations.

2. Methods

TransPipe was developed to manage the translation and transcription process that was, previously, a combination of online tools and manual work without medical or WHO-specific terminology banks. OpenWHO conducted six iterations to test the transcription and translation accuracy with the new external solution provider, and to fine-tune the

workflow with TransPipe. Feedback generated from testing was used regularly for improvements throughout the project. As mentioned before, TransPipe itself is not the actual service provider, but the connection between the learning platform and several external services. TransPipe integrates with OpenWHO through a custom Application Programming Interface (API) and thereby has access to all course videos of the learning platform. Upon request, TransPipe uses the following two main services offered by external providers to process the videos and create the subtitles:

1. automated speech-to-text transformation to generate video subtitles in the original language,
2. automated translation from one language to another to additionally provide video subtitles in additional languages.

During the implementation of TransPipe, WHO and HPI collaborated with service providers to improve the quality of their services. At a certain point, WHO focused on one of these service providers and tested the accuracy of transcribing English audio from OpenWHO course videos and the subsequent transcript translation into the UN's official languages (Arabic, Chinese, French, Russian, and Spanish) and Portuguese. The tests highlighted differences in accuracy for different languages, with Latin script languages (French, Spanish, and Portuguese) being translated with 70-90% accuracy. At the same time, testers found Arabic, Chinese, and Russian translations to be around 60% accurate. However, subsequent tests found improved accuracy across the board as the default terminology and vocabulary bank expanded with the tool's repeated use. At the upper end of the spectrum, the translation accuracy rivaled competitor transcription tools. Accuracy could continue to be improved with the development of a custom terminology bank populated with WHO-specific terminology and vocabulary, which was impossible with existing competitors at the time.

Once WHO had ascertained a satisfactory accuracy level through internal testing, OpenWHO's platform service provider, the Hasso Plattner Institute (HPI), included the selected external transcription and translation solution in TransPipe's list of supported services. The selected external solution also includes a terminology tab to enable users to identify the errors caused by the machine and save the correct terminology in the repository for future use. Connecting TransPipe to the OpenWHO platform via a dedicated API finalized the deep integration of the new process within OpenWHO's workflow. The ultimate goal was to provide learning materials in different languages, achieving sufficient accuracy to reduce the turnaround time on each piece of content from a few days to a couple of hours. The newly created TransPipe tool also improves collaboration within a team of language experts on the subtitles. It allows assigning specific team members to focus on particular languages of the videos and can automatically notify others when the workflow may proceed. Today, TransPipe facilitates the transcription and translation of all OpenWHO's transcripts and subtitles.

3. Results

By connecting the existing services and filling in the gaps, TransPipe has fully replaced the different manual steps of the previous workflow with automated or semi-automated processes. Particularly, it has eliminated the need to use error-prone workflows including desktop tools, and provides a uniform layer atop of all transcription and translation services that have been used before. Because of this, it has become a one-stop shop for course administrators and editors (translators and QA managers), who can now all work

with the same web-based application. The languages in which the course is to be translated, and the required transcription and translation services can now be configured. Then TransPipe triggers the transcription by invoking the APIs of the configured transcription service, and translators and QA managers can be appointed for each selected language.

Translation for the other languages can be triggered. Editors are notified upon completion via email, which allows them to log into the TransPipe interface and start the quality control process. TransPipe always provides a visualization of the video and course status as all transcription and translation tools offer asynchronous processes to allow some time to return the results. The same applies for the human QA processes. Once the authorized person in charge decides that the quality is sufficient, one more click is needed to publish the subtitles in all translations for all course videos. Figure 2 compares the subtitle generation process with and without TransPipe.

As TransPipe was rolled out, it enabled a process for publishing fully automated multilingual subtitles on OpenWHO, requiring very little manual human input. By the end of 2022, the TransPipe tool will have contributed to the transcription of nearly 4,700 minutes of video content and the translation of 1,050,700 characters in the context of video subtitles on OpenWHO. As of November 2022, TransPipe has primarily been used to transcribe a total of 4,250 minutes of English audio whereas Arabic has amounted to some 100 minutes.

TransPipe has been used to translate English subtitles into other official UN languages, such as Arabic, French, Spanish and Russian, as well as Italian, Portuguese, and Turkish. In addition, it is possible to use TransPipe to translate into seven other target languages (18 languages in total), including some for which the reliable acquisition of translators has proven difficult, such as Telugu and Tamil. The connected external service offers an even more significant number of source and target languages (18 and 51, respectively), which WHO and HPI could enable on TransPipe in the future. Being powered by machine learning, the translation provision integrated with TransPipe continuously augments its vocabulary bank to improve transcription and translation accuracy.

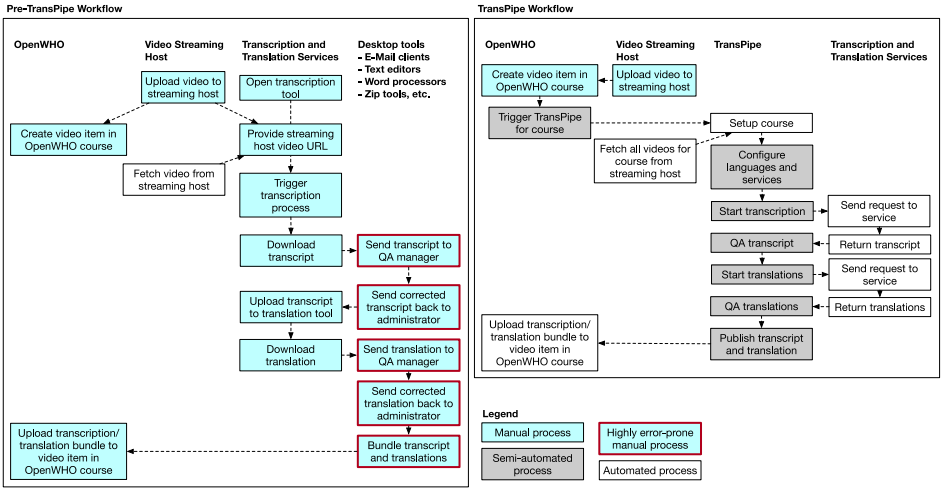


Figure 2. Pre-TransPipe vs. TransPipe process

4. Discussion

Automated transcription and translation have enormous potential as public health learning tools. It has allowed the near-simultaneous availability of video subtitles on OpenWHO in many languages. In addition, the time saved by automatic transcription and translation can hardly be understated compared to doing the same manually. However, there are also limitations. For example, for under-resourced languages, in particular, the validation process could still be challenging despite the improved efficiency [4]. The missing evaluation of the linguistic complexity in the source language also raises questions about quality consistency and learning standards. Continuous research and effort are required to fully realize its potential, including the continued creation of tailored training datasets for optimal transcription and translation accuracy and solidifying a community of practice for testing and rollout of novel tools. Once these tools are well-established and provide satisfying results, a seamless integration into existing contexts and applications is required to leverage the full potential. With the work at hand, we perform the first step in this direction. However, further questions arise, for example regarding the maintenance of user-defined terminology across individual videos or the use of the manually corrected transcripts and translations for supervised machine learning.

5. Conclusion

Technology must serve humanity and keep pace with our needs. The Covid-19 pandemic has provided unprecedented opportunities to innovate and make the fruits of technology available to increasingly large numbers of people. The principle of equity is essential in all work to achieve the broadest of public health goals, especially in health emergencies. For WHO, the OpenWHO platform and technological advances must help remove barriers such as cost, language and digital access so that anyone anywhere can access life-saving knowledge.

References

[1] Che X, Luo S, Yang H, Meinel C. Automatic Lecture Subtitle Generation and How It Helps. ICALT: IEEE 17th International Conference on Advanced Learning Technologies. 2017; 34-38, doi: 10.1109/ICALT.2017.11.

[2] Zhao U, Samo G, Utunen H, Gamhewage G, Stucke O. Evaluating Complexity of Digital Learning in a Multilingual Context: A Cross-Linguistic Study on WHO's Emergency Learning Platform. ICIMTH: Studies in Health Technology and Informatics. 2021; 281:516-517, doi: 10.3233/SHTI210222 2021.

[3] Bethge J, Serth S, Staubitz T, Wuttke T, Nordemann O, Das PP, Meinel C. TransPipe: A Pipeline for Automated Transcription and Translation of Videos. EMOOCs: 7th European MOOC Stakeholder Summit. 2021; 79-94, doi: 10.25932/publishup-51030.

[4] Besacier L, Barnard E, Karpov A, Schultz T. Automatic speech recognition for under-resourced languages: A survey. Speech Communication. 2014; 56:85-100, doi: 10.1016/j.specom.2013.07.008.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230163

Extended Reality Is Underutilized in Medical Device Training: A Descriptive Literature Review

Minna LAUKKAVIRTA[a,b,c,1], Krista TARKKALA[b,d], Marjo TIENARI[e] and Tuija S. IKONEN [b,c]
[a] *Kanta-Häme Central Hospital, Hämeenlinna, Finland*
[b] *University of Turku, Department of Public Health, Turku, Finland*
[c] *Ostrobothnia Wellbeing Services County, Finnish Centre for Client and Patient Safety, Vaasa, Finland*
[d] *Helsinki University Hospital, Hyvinkää, Finland*
[e] *Tampere University of applied sciences, Tampere, Finland*
ORCiD ID: Minna Laukkavirta https://orcid.org/0000-0002-6475-6136

Abstract. This descriptive review provides a synthesis of existing literature about the use of extended reality (XR) including virtual (VR) and augmented reality (AR) technology solutions for competence assurance, training and orientation regarding digital skills and medical device training. From the literature, only few original studies were recognized with a study question or aim to assess medical device training as the target of virtual training modalities. XR methods could provide potential useful solutions to improve medical device competence. Based on the literature, it was evident that further studies are required to research the possibilities of XR technologies to improve medical device training.

Keywords. Medical device, patient safety, augmented reality, virtual reality

1. Introduction

Safe and competent use of medical devices in patient care is of paramount importance to both the patient and the medical staff [1, 2]. Despite a clear need for training and orientation to improve medical device safety, there are major challenges how to acquire and provide sufficient education [3]. In Europe, the European medical device legislation has been renewed and the Medical Devices Regulation (Regulation (EU) 2017/745) has been in effect since 26 May 2021. In Finland, the legislation concerning medical devices (Finlex 629/2010) defines the responsibilities of the professional user, as well as the responsibilities of the employer to ensure adequate competences of the staff. Medical device training and orientation need to be more effective in order to ensure a sufficient level of health technology and medical device skills among health care personnel. Since training and orientation is challenging during working hours, it is necessary to develop new types of pedagogical extended reality (XR) solutions including virtual (VR) and

[1] Corresponding Author: MD, PhD Minna Laukkavirta, Matarakaari 26 b, 21500 Piikkiö, Finland, E-mail: minna.laukkavirta@iki.fi.

augmented reality (AR) technology for digital and medical device competence assurance, training, and orientation. Aim of this study was to assess how extended reality technology has been utilized in medical device training to enhance learning and patient safety.

2. Methods

As part of the Virtual Platform for Medical Device Training project, the published scientific literature was analysed in order to describe the status quo of the XR solutions for training and learning health care technologies. The aim was to identify studies focused on medical device safety in health care, including hospital based, home care, and social welfare sector. The purpose of this review was descriptive.

The literature search was performed initially between October and November 2021 and a complementary search was performed in March 2022. The search was performed using Medline, PubMed and Google Scholar and Google. In addition, searches were completed from the reference lists of the detected articles. There was no time limitation. English, German, French and Scandinavian languages were included.

One researcher (TI) went through the relevant headings and abstracts and selected the full text articles for reading to other members of the review team. Each member was allocated a subheading for analysis of the literature. The focus was on the effect of immersed technologies on learning of skills required for safe use of medical devices.

The following search phrases were used in various combinations: virtual reality, augmented reality, extended reality, teaching, education, learning, medical device, medical device safety, medical device training, learning of medical device safety, medical equipment, medical device safety learning, medical device training simulation, review.

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart was used, although the study did not fill all other PRISMA requirements.

3. Results

The search protocols and the number of abstracts and full text articles are described in Figure 1. Altogether 1375 headings, and of those 48 abstracts of interest were found. The authors performed complementary searches while writing the descriptions. Four articles were added to the review material as result of complementary search during writing the review. In total 31 full text articles and two theses were read and five abstracts of interest were included even though full text was not available.

The review confirmed that extended reality training modalities are already being used as a part of health care education in nursing, medical, dental, and paramedic education. Simulation offers a safe environment for training different skills and for learning from mistakes. The attitudes towards training are mainly positive in the field on health care. Simulated authentic patient situation is considered the most motivating training [4].

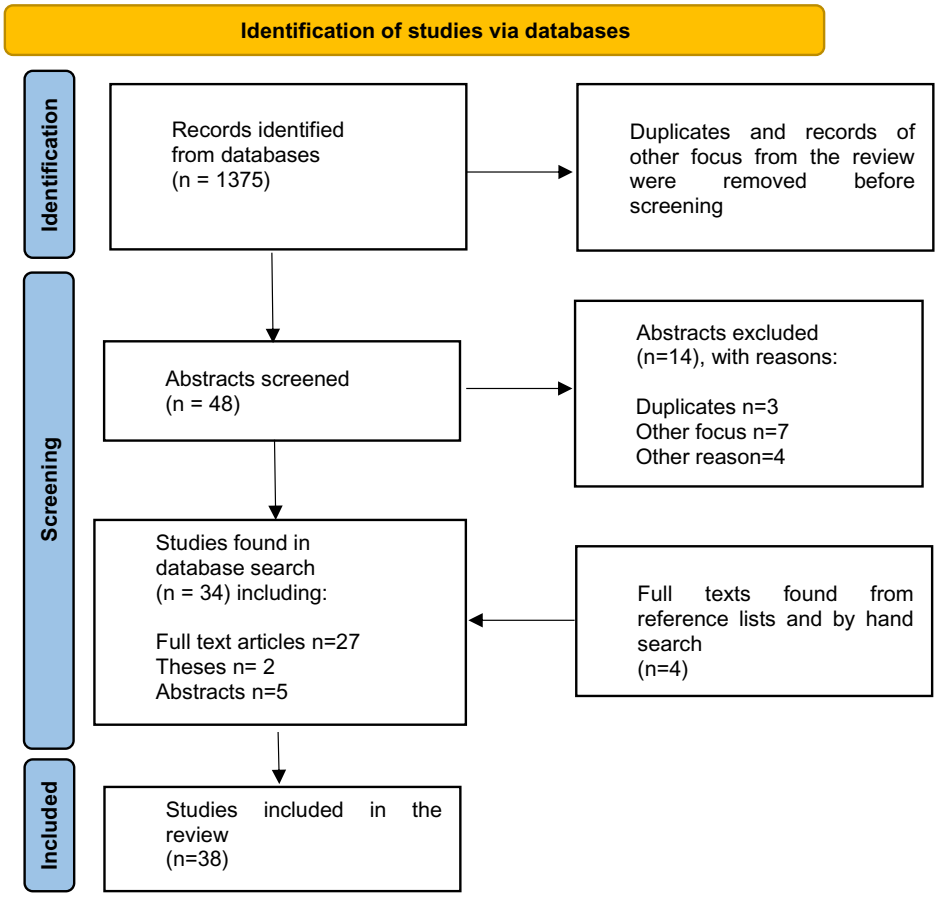


Figure 1. PRISMA flow chart for presenting the selection of literature.

Typically, medical or surgical procedures formed the majority of trained techniques, such as neurosurgery, gastrointestinal surgery, orthopedic surgery, dental surgery, surgical ophthalmoscopy, gynecology, surgical knot training, and peg tube, central line and catheter insertion. Anatomy teaching and emergency medicine decision-making skills were trained, as well.

VR technology was also being used in the fields of leadership, communication, decision making, critical thinking, inclusivity, health appraisal and disaster triage [5]. VR assisted learning methods were more effective than the control conditions in improving knowledge, but there was no difference between skills, satisfaction, confidence, or performance time [5].

It was concluded that online simulation training provides a virtual hands-on experience in advance of implementing new equipment and coupled with adult learning

principles, may enhance the experience of clinical end users of technology. Especially in home environment, augmented technology will add skills to manage safety hazards [6].

Although VR technology is developing rapidly, the immaturity on the technology can cause challenges to the students' ability to transfer "hands-on skills" from VR environment to reality [5]. AR seems to be more effective in supporting skill development rather than knowledge gain when compared to other techniques [7].

Immersive experiences may foster the teaching and learning of complex medical contents. Head mounted devices (HMD) including a headset or glasses are versatile, low-price, and mobile, and make learning content more accessible and engaging [8, 9]. The training with HMDs were shown to improve practical skills, reduce stress, and gain self-confidence for the actual procedure. The main benefits of studies have shown decreased surgical error rates, cost-effectiveness, and improved knowledge [8, 10]. VR training also increased confidence and reduced anxiety, and training improved measured understanding, technical skills and immersed efficiency [11].

Not all studies reported effective outcomes for the use of HMDs in medical education. The disadvantages of HMDs were motion sickness and nausea, technical problems, and stress. These may impede learning and training. It is unclear if symptoms of motion sickness and nausea are related to beginners or if they tend to persist and affect learning. Women more often suffer from motion sickness using VR devices. As AR devices combine real and virtual environments, they seem to mitigate negative health effects such as blurred vision, disorientation, and cybersickness [8, 10].

Although improved medical device skills are linked to patient safety, only one study described training of safety during medical device usage as the main purpose of the study. The use of virtual training for fire prevention was investigated while using electrocautery device. The authors concluded that training for operating room fire emergencies in fully immersive VR environments may be the ideal training modality [12].

4. Discussion

Based on the current literature, only few original studies were recognized with a study question or aim to assess medical device training as such, as a target of extended reality training modalities. Majority of the studies were not about how to use the medical device, but were simulations of procedures and about learning operative or procedure skills.

The teaching methods based on the utilization of XR in health care education are used somewhat integrated with more traditional teaching methods. The use of virtual technologies is increasing and being developed to be more user friendly. Cost efficiency and relative independence from time and place are being identified as advantages of using extended reality in learning. Such technologies offer the possibility of scalability and repeated practice without adverse effects on the patient.

The majority of studies included in this review considered extended reality based intervention as at least non-inferior to the traditional teaching methods. Even though such training methods have been reported as an engaging and enjoyable tool for learners to improve their knowledge and skills, a considerable proportion of users suffer from physical adverse effects of the technology that might impede learning. Usability and availability of the technology and the adverse effects experienced by the user are seen as the challenges of using virtual technologies in teaching.

The limitations of the review include some shortcomings of the full commitment to PRISMA checklist. For example, the risk of bias was not assessed for the selected articles,

and the numbers of participants were not counted. Apart from searched databases, there might be other databases containing virtual training publications. Due to time limitations, only one researcher read the headings and abstracts. Each author did the analysis and writing mainly independently, which might be seen as a limitation to the review.

5. Conclusion

The reviewed original studies posed positive expectations to extended virtual technologies as a solution for educational and training requirements concerning medical devices. Future studies are required that focus on teaching the safe use of health technologies in order to maximize patient safety. The benefits of XR solutions in learning medical device safety have not yet sufficiently researched. Thus, more studies are needed.

Acknowledgement. The Virtual Platform for Medical Device Training project has been funded from REACT-EU instrument as part of the European Union COVID-19 pandemic recovery measures.

References

[1] Brand D, Just a piece of equipment? The importance of medical device education. J Perioper Pract, vol. 22, no. 12, pp. 380-2, Dec 2012.
[2] Shields R and Latter K, Improving the uptake of medical device training to promote patient safety. Nurs Stand, vol. 35, no. 1, pp. 31-34, 2019.
[3] Swayze SC and Rich SE, Promoting safe use of medical devices.,Online J Issues Nurs, vol. 17, no. 1, p. 9, 2011.
[4] Mäkinen H et al, User experiences of virtual reality technologies for healthcare inlearning: an integrative review. Behaviour and Information Technology, Volume 41, Issue 1, 2022.
[5] Chen FQ et al., Effectiveness of Virtual Reality in Nursing Education: Meta-Analysis. J Med Internet Res, vol. 22, no. 9, p. e18290, 2020.
[6] Devers V, Use of Simulation-Based Training to Aid in Implementing Complex Health Technology. Biomed Instrum Technol, vol. 52, no. 1, pp. 44-48, 2018.
[7] Parsons D and MacCallum K, Current Perspectives on Augmented Reality in Medical Education: Applications, Affordances and Limitations. Adv Med Educ Pract, vol. 12, pp. 77-91, 2021.
[8] Barteit S et al, Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. JMIR Serious Games, vol. 9, no. 3, p. e29080, 2021.
[9] Zhu E et al., Augmented reality in healthcare education: an integrative review. PeerJ, vol. 2, p. e469, 2014.
[10] Jensen L and Konradssen F, A review of the use of virtual reality head-mounted displays in education and training. Education and Information Technologies 23, 1515-1529, 2018.
[11] Edwards TC et al., Immersive virtual reality enables technical skill acquisition for scrub nurses in complex revision total knee arthroplasty.Arch Orthop Trauma Surg, vol. 141, no. 12, pp. 2313-2321, 2021.
[12] Dorozhkin D et al., OR fire virtual training simulator: design and face validity. Surg Endosc, vol. 31, no. 9, pp. 3527-3533, 2017.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230164

Informing Family About Patient Trajectory During Surgery: Design and Preliminary Evaluation

Frederic EHRLER[a,1] and Alexandre JELEFF[b]
[a] *University Hospitals of Geneva, Switzerland*
[b] *Department of Anesthesiology, Pharmacology, Intensive Care and Emergency Medicine, University Hospital of Geneva*

Abstract. The patient empowerment movement has highlighted the importance of providing information to the patients to improve care outcome. However, relatives of patients are not yet taken into consideration. This is especially problematic during surgeries since families are often left without real-time information about the trajectory of the patient, inducing worries. Based on this observation we have developed the SMS-Chir solution that connects our surgery service management system with the automatic sending of SMS at key moments in order to inform families about the progression of the surgery. The system has been conceived thanks to the results of a focus group involving four experts. The evaluation was done by monitoring the use of the system over time and by sending questionnaires after intervention. Results analysis shows a limited use of the system but a high satisfaction of the beneficiaries. This study highlights the importance of managerial factors (resistance to change) in order to onboard the necessary stakeholders in the process.

Keywords. Surgery, Patient empowerment, Communication

1. Introduction

The healthcare system is becoming ever more patient-focused, with an increasing emphasis on patient empowerment. Patient empowerment is a concept that encompasses the ability of individuals to take control of their own health, to make informed decisions, and to actively contribute to their healthcare management. This trend, towards patient empowerment, is driven by the benefits it offers, such as improved health outcomes, better use of healthcare resources, and more personalized healthcare. However, there is a lack of attention to the needs of family members of patients, particularly in the context of medical procedures that may induce significant anxiety and stress. This is especially true for surgeries, which can be a highly stressful and anxiety-provoking experience for family members due to the lack of information about the patient's trajectory, from preoperative assessment to postoperative recovery. As a result, when no information is provided, families tend to harass caregivers to obtain information about the progression of the surgery. The request for information do not only interrupt frequently caregivers

[1] Corresponding Author, Frederic Ehrler, University Hospitals of Geneva, Geneva, Switzerland; E-mail: Frederic.ehrler@hcuge.ch.

but also creates additional frustration each time caregivers do not have information to provide.

In response to this issue, a new system has been designed to address the lack of communication between healthcare professionals and family members [1]. This system consists of textual messages sent automatically to the family member's mobile phone at predetermined moments of the patient's care trajectory, such as preoperative assessment with anesthesiologists, admission to the operating theater and postoperative recovery [2]. The messages provide an update on the patient's progress and enable the family member to stay informed. This system has the potential to improve the experience of family members of patients undergoing a surgery and to reduce their levels of anxiety and stress [3].

The aim of this article is to discuss the implementation and evaluation of this new system. We will attempt to understand the importance of providing family members with meaningful and timely information as well as whether the information provided can be a source of stress for family members.

2. Method

2.1. Need analysis

The need analysis for our study was conducted by interviewing four experts from the operating theater, surgery programming, nurse directorate and IT department during a focus group. The experts were selected based on their expertise in the design of a system aiming at informing family about the progression in surgery of a patient. The focus group was conducted in a semi-structured manner [4]. We prepared a list of questions concerning the design of the system, such as what are the relevant moments to send information? who are the eligible patient? Followed by a more general discussion around the potential challenges that could arise when implementing the system. The questions were designed to elicit the opinions of the experts and to spark discussion. The experts were given the opportunity to provide additional comments or to ask questions.

2.2. Evaluation of the system

For the evaluation we relied on a descriptive assessment of the effectiveness of the SMS system. We included as participants all family members of patients who underwent surgery at a single medical center and were willing to use the system. The data were collected through a questionnaire whose link was provided by an additional SMS automatically sent after using the system. Participants were asked to complete the survey assessing the clarity of the information, its usefulness, and the reduction of anxiety. Participation was voluntary and anonymous. We analyzed the data using descriptive statistics.

3. Results

3.1. Design of the system

During two focus group sessions, four experts from the operating theater, surgery programming, nurse directorate and IT department were brought together to discuss the solution.

One of the strong constraints identified during the focus group was the need to rely on the events already available in the system of surgical theater management. The HUG relies on the Centricity Opera software developed by GE that is used to plan as well as monitor the use of resources in operating theatre. This software allows to monitor the surgery progression based on events reported by different actors all along the care process. In total there are ten different events that are recorded in the system that can potentially trigger a message to the family. These events range from the arrival in the surgical theater to the return of the patient in the care unit. Choosing the relevant events to communicate with the family was not straightforward. Indeed, if a strong consensus was found to trigger a message at the beginning of the surgery, the discussions were livelier for the other events. Indeed, for the end of surgery it is not unlikely that the event linked to the end of surgery is not reported directly in the system because the person in charge must prepare the operating theater for the next patients and may postpone the documentation task. The consequence would be that the SMS is delayed creating anxiety for the family. Therefore, it was necessary to ensure good communication among the surgery personnel to ensure the timely report of the end of surgery event. Finally, the last SMS sent is the questionnaire to the family to evaluate the satisfaction.

The second step was to define the content of the messages to send to the family. Choosing the SMS content had to be done carefully in order to respond to two constraints, the length of the SMS and the respect of privacy. Indeed, it was necessary to keep the message short to avoid costs associated with the sending of SMS. Regarding privacy, the messages must transit through international servers therefore the information transmitted should remain anonymous. After validation by the medical director and the jurist the following messages were implemented. *Hello, we inform you that the surgery of Mr/Mrs Surname Name first letter has started at HH.mm.*

3.2. Evaluation of the system

From September 2021 to December 2022, we monitored the use of the system that was open to every patient from the thoracic surgery. 117 patients were enrolled in the system during this period which represent 15.8% of the overall eligible patients (741).

We observed that the use of the system remained limited over time, if at the start of the project the trend seemed to increase quickly, the frequency of use dropped after that and then varies over the months without increasing. On this period only 7 events were missing at the start of the surgery. Regarding the satisfaction evaluation 35 (30.8% of the participants) person replied to the satisfaction questionnaire.

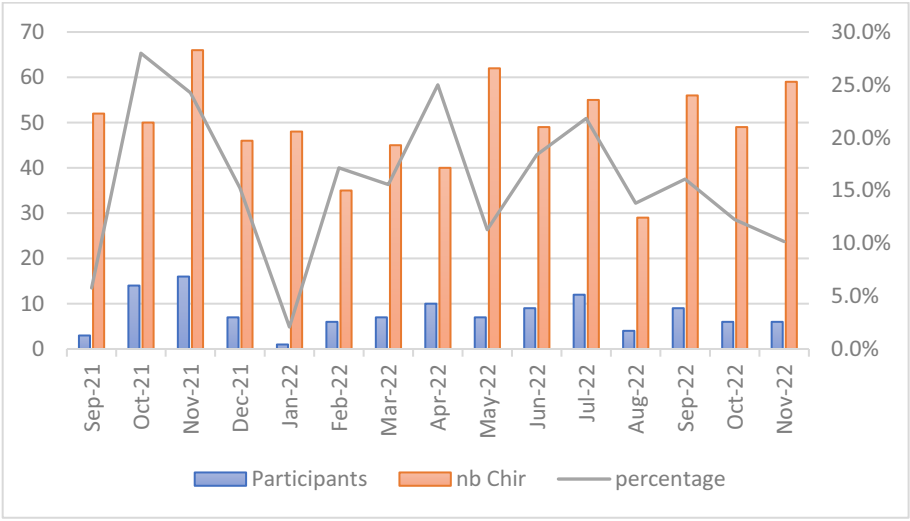


Figure 1. Use of the system between September 2021 and November 2022.

Table 1. Results of the satisfaction questionnaire

	strongly agree	Agree	disagree	Strongly disagree
Satisfaction	23 (65.7%)	11 (31.4%)	1 (2.9%)	0
Anxiety reduction	12 (34.3%)	21 (60.0%)	2 (5.7%)	0
Need for additional info	23 (65.7%)	10 (28.6%)	2 (5.7%)	0

Participants were strongly satisfied with less than 3% of the patients unsatisfied. Participants were also questioned whether they would like to receive information at additional moments of the patient trajectory. 24 (66.7%) would like to know when the patient returns to the care unit, 19 (54.3%) would like to know when the patient arrives in recovery room as well as departure from it.

4. Discussion

4.1. Managing the uncertainty

Surgery puts patients' health at risk due to many factors such as patient frailty, anesthesia, and complexity of the procedure. For patients with high risk factors there is no guarantee that the surgery will unfold as expected. It is not unlikely that complications are encountered during the surgery impacting its duration. In more severe situations, a patient may require intensive care or even die during the procedure. In such a case, it raises ethical question whether it is acceptable to automatically provide information to the family without being accompanied by explanation [5]. Thus, we recommend that caregivers should remain central in the relationship with the patient [6].

4.2. Involving stakeholder

One of the important difficulties encountered during the project is the enrollment of the family in the process. Indeed, the service must be proposed by the personnel in charge of programming surgery schedule. The employee enrolling the patient is perceived this task as an additional workload, whereas its benefits go to the caregivers in the care unit that are less disrupted by frequent requests by the families. The importance of perceived benefits has been demonstrated as an important dimension of technology acceptance. Also, another dimension that should be considered is the support from the hierarchy that must give clear instruction and objectives to the persons in charge [7].

5. Conclusion

Although patient empowerment interventions have become common in healthcare, the families of patients remain still discredited and lack of information about the situation of their relatives. This problem is particularly acute in surgery since patient's health is at risk and information to the families can be sparse, inducing a high level of stress. Following the implementation of the system, we observe that the system lead to high satisfaction confirming the findings of similar system in other contexts [1–3]. There is however an underutilization due to the resistance of surgery personnel and anesthesiologists to propose the service. This lack of motivation certainly due to misperception of the benefits can be a barrier to its full potential. Among the possible solutions, we could communicate about the benefits to the patients or to empower the patients one step further by letting them choose autonomously, through a patient portal, who are the people they want to inform.

References

[1] Kim S, Jeong D, Ryu MH, Sok SR. Effects of Information Delivery Intervention for the Family of Patients Undergoing Surgery on Emergency in South Korea. SAGE Open. 2021 May;11(2):21582440211016555.. doi:10.1177/21582440211016555.

[2] Howe LS, Wigmore D, Nelms N, Schottel P, Bartlett C, Halsey D, Krag M, Lunardini D, Monsey R, Beynnon B, Blankstein M. Perioperative family updates reduce anxiety and improve satisfaction: a randomized controlled trial. Journal of Patient-Centered Research and Reviews. 2021;8(2):107.

[3] Mignault A, Nguemeleu ÉT, Robins S, Maillet É, Matetsa E, Dupuis S. Correction: Automated Intraoperative Short Messaging Service Updates: Quality Improvement Initiative to Relieve Caregivers' Worries. JMIR Perioperative Medicine. 2022 Jan;5(1). doi:10.2196/36208.

[4] Kitzinger J. Qualitative research: introducing focus groups. Bmj. 1995 Jul 29;311(7000):299-302. doi:10.1136/bmj.311.7000.299.

[5] Pérez-Cárceles MD, Pereniguez JE, Osuna E, Luna A. Balancing confidentiality and the information provided to families of patients in primary care. Journal of medical ethics. 2005 Sep 1;31(9):531-5.

[6] Fuentes-Ramirez A, Laverde-Contreras OL. Nursing intervention to meet the family members' needs during the surgery waiting time. Revista Latino-Americana de Enfermagem. 2021 Oct 29;29. doi:10.1590/1518-8345.5028.3483.

[7] AlQudah AA, Al-Emran M, Shaalan K. Technology acceptance in healthcare: A systematic review. Applied Sciences. 2021 Jan;11(22):10537. doi:10.3390/app112210537.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230165

Evidence of Digital Health Applications from a State-Regulated Repository for Reimbursable Health Applications in Germany

Urs-Vito ALBRECHT[a,1], Ute VON JAN[b], Dennis LAWIN[a,c], Evgenii PUSTOZEROV[a] and Florian DITTRICH[a,d]
[a]*Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Bielefeld, Germany*
[b]*Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
[c]*Department of Cardiology and Intensive Care Medicine, University Hospital OWL of Bielefeld University, Campus Klinikum Bielefeld, Bielefeld, Germany*
[d]*Joint Centre Bergisch Land, Department for Orthopaedics, Sana Fabricius Clinic Remscheid, Remscheid, Germany*
ORCiD ID: Urs-Vito Albrecht https://orcid.org/0000-0002-8989-6696, Ute von Jan https://orcid.org/0000-0001-9225-593X,
Dennis Lawin https://orcid.org/0000-0002-1292-7845,
Evgenii Pustozerov https://orcid.org/0000-0001-8109-1319,
Florian Dietrich https://orcid.org/0000-0002-5135-4736

Abstract. 17 RCTs for 15 digital health applications (DiGA) permanently listed in the state-regulated register were analyzed descriptively for methodological study aspects relevant to evidence analysis. The analysis revealed that several underlying studies had limitations, at least worthy of discussion, in terms of their power concerning sample size, intervention and control group specifications, drop-out rates, and blinding.

Keywords. health apps, evidence, reimbursement, digital health application

1. Introduction

In Germany, the Digital Health Care Act (DVG), passed in November 2019, enables doctors and psychotherapists to prescribe "DiGA" (digital health applications) that are covered for those enrolled in statutory health insurance, based on criteria specified in the DVG [1] and the Digital Health Application Regulation (DiGAV) [2]. DiGA are defined as digital, certified low-risk medical devices that help "[...] to support the detection, monitoring, treatment or alleviation of diseases or the detection, treatment, alleviation or

[1] Corresponding Author: Urs-Vito Albrecht, Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany; E-mail: urs-vito.albrecht@uni-bielefeld.de.

compensation of injuries or disabilities [...]" [1]. This covers not only apps but also browser-based applications. However, only applications listed in the directory of digital health applications (DiGA-Verzeichnis, DiGA-VZ) at the German Federal Office for Drugs and Medical Devices (BfArM) [3], after successful completion of an assessment, can be prescribed by physicians or therapists or reimbursed after approval by the health insurer. Permanent inclusion in this directory occurs only if an application has successfully demonstrated interoperability, adequate consideration of data protection and data security, meets the requirements for medical device status (safety, functionality, quality), and if it has demonstrated a "positive impact on care" (PIC). In the absence of sufficient evidence regarding the positive impact of an app, it is possible to provisionally include the app in the directory for a limited period (12 months, extension to 24 months max. is possible), during which time corresponding evidence has to be provided. In this article, the authors provide descriptive data for permanently listed applications and discuss the acquired data under evidence aspects.

2. Method

Available meta-tagged information was extracted from the web-based interface provided by BfArM (DiGA-Verzeichnis, [3]) on December 5th, 2022. The data was stored in tabular form and initially evaluated with respect to the apps' listing status (i.e., provisional or permanent). Entries with a permanent status were subjected to further descriptive analysis (counts, percentages (%), mean values (m), standard deviations (sd), median (md), interquartile range (iqr)) with respect to indication, primary endpoints, control group, and setting, the number of participants, percentage of drop-outs, duration of use and follow-up within the study context, blinding status, National Institute for Health and Care Excellence (NICE) classification [4], and Agency for Healthcare Research and Quality (AHRQ) criteria [5,6].

3. Results

The DiGA directory [3] contained 15 permanently listed applications. Prices ranged between 189€ and 599€ (m: 333,43€; sd: 169,68€). 4 DiGA were available natively for iOS and Android, and there were 8 purely web-based apps and 3 with both web and native implementations. 9 DiGA were dedicated to the BfArM category "psyche," 2 DiGA to "hormones and metabolism," and 1 DiGA each was assigned to the categories "muscles, bones and joints," "nervous system", "ears," and "other". A listing of the indications addressed by the applications can be found in Table 1. For all permanently listed apps, published evidence has been deemed to sufficiently support medical benefit and efficacy according to the BfArM's requirements. However, there are not always links to peer-reviewed publications in this regard (see the cited publications and footnotes in Table 1). Also, all DiGA fully meet the corresponding evidence requirements of the NICE category 3b (i.e., therapeutic purposes, meaning the app "provides treatment for a diagnosed condition or guides treatment decisions"), independent of BfArM logic ("Evidence of positive benefit-risk ratio by valid comparative studies or at least one RCT"). 14 of these apps have an AHRQ evidence level of "Ib" ("At least one sufficiently large, methodologically high-quality RCT"), and for one app, there is higher level

evidence at level "Ia" ("At least one meta-analysis based on methodologically high-quality randomized controlled trials").

Table 1. Methodological study aspects of 15 permanently registered DiGA.

No	Primary Endpoint	Ctrl.	Part.[n]	DrpOut I; C[%]	Use	FlwUp[m]	Bld	AHQR
01a[7]	Depression	std	1013	26; 25	3	6	yes[a]	Ib
01b[8]	Depression	std	163	27; 28	3	6	no	Ib
01c[9]	Depression	mixed	2901	– [b]; – [b]	2–3	Mixed	no[c]	Ia
02[10]	MS related fatigue	std	275	32; 15	3	24	no	Ib
03[11]	depression (diabetics)	online education	260	24; 12	2	– [d]	no	Ib
04[12],e	Panic	wait list	92	22; 21	2	6	yes[f]	Ib
05[13]	Stress	wait list	264	30; 8[g]	1.75	12	no	Ib
06[14],h	Vaginism	wait list	200	–; – [i]	3	6	no	Ib
07–[h]	phobia/panic	wait list + support	297	18–21; 6–19[j]	2	6	no	Ib
08[15],h	Tinnitus	std	187	–; – [k]	3	9/12[l]	no	Ib
09[16],m	Depression	wait list	401	57-59[n]; 71	3[o]	6	yes[p]	Ib
10–[h]	Anxiety	wait list	156	27; 13	3	3	no	Ib
11[17]	Insomnia	wait list	56	10; 0	1.5	12[q]	no	Ib
12[18]	Anxiety	std	139	20; 13	3	6	no	Ib
13[19],r	back pain	std	215	–; 2	10	3	no	1b
14[20]	alcohol consumption	std	608	37; 23	6	6	no	Ib
15[21],s	obesity	cont. as usual	149	7[t]	9	12	no	Ib

All 17 corresponding studies were prospective randomized controlled trials (RCT) following the Intention-to-Treat principle. For one application (01), there was also a meta-analysis of multiple RCTs related to the app. Legend: "Ctrl.": Control group; "Part.[n]": number of participants; "DrpOut I;C[%]": Dropout rates for intervention and control group; "Use, FlwUp[m]": use and follow up periods in months; "Bld": blinding; std: standard care

[a] Assessor-blinded.

[b] 12 studies were included, with dropout rates between 6 and 56%.

[c] Assessor blinding was impossible, as only studies with self-reported assessments were included.

[d] While the study (available with study id DRKS00004748 on https://drks.de/) is described as using a 6 and 12-month follow-up assessment, the publication [11] does not provide data related to this.

[e] Publication listed on https://hellobetter.de/online-kurse/panik/.

[f] Interviewers were blinded to the participant's randomization status.

[g] Data reported at 12 months for intervention (dropout at 6 months: 12.9%), 6 months for the control group.

[h] References to peer-reviewed publications lacking in the DiGA directory, but specified on the DRKS homepage https://drks.de/ (App 06: DRKS00010228, App 08: DRKS00022973, App 10: DRKS00023799) or on https://clinicaltrials.gov/: App 07: NCT0551080).

[i] Dropout data pay-walled; the abstract states: "on average, participants completed 79% of the intervention."

[j] Values/diagnosis: Agoraphobia, I: 18%, C: 6%); panic dis., I: 21%, C: 12%; social phob., I: 21%, C: 19%.

[k] Dropout values not stated within the provided data [15].

[l] 9 months for intervention, 12 months for the control group.

[m] Publication identified from the manufacturer's homepage.

[n] Dropout for follow-up, two intervention groups (guided: 57%, unguided: 59%).

[o] Intervention was available after the initial 12 weeks until follow-up.

[p] "Interviewers were blinded to the assigned group of individuals" [16].

[q] Intervention group only.

[r] No peer-reviewed publication for the RCT is listed in the directory, however, recently stated on the manufacturer's homepage.

[s] The preliminary RCT data (not yet published in a peer-reviewed journal) not specified in the DiGA directory, but identified via PubMed, describes the use by approx. 11,000 users in Germany. Data shown here were, however, obtained from the DiGA directory.

[t] Value for intervention and control group combined (at 9 months).

4. Discussion

In addition to fulfilling all technical requirements of the DiGA, it is also crucial to evaluate the medical benefit by providing evidence for PIC. By creating this new definition, the BfArM has given its concept of quality a framework in terms of terminology. However, the prerequisite for this is submitting a scientific evaluation concept prepared by a "manufacturer-independent institution to prove PIC, as well as the medical services required for testing" [1]. All 15 permanently listed applications in the German DiGA repository show at least an evidence level of Ib (AHRQ) and a NICE category of 3b. However, there are methodological limitations among the underlying RCTs to some degree, and thus they carry a higher risk of bias. For example, the number of study participants in the RTCs is quite small compared to numbers commonly included in clinical trials (md: 215; iqr: 141), which raises questions about the validity of the studies. The frequently higher dropout rates in the intervention groups (md: 26%; iqr: 10%) compared with the control groups (md: 13%; iqr: 15%) may indicate problems with the internal validity of some RCTs. Possible systematic errors may be responsible for these dropout rates. In most cases, the definition of the "standard of care" (SOC) applied by the investigators is ambiguous. Therefore, the interventions' advantages over SOC are not amenable to interpretation. Overall, the study endpoints are often only imprecisely aligned with the intervention objectives. Also, it is often unclear whether participants assigned to the control groups were given any previous treatment or were naive to treatment before randomization. Another factor for bias is the lack of blinding against interventions: only 3 of the 17 publications we evaluated were blinded. Quality control is debatable for 3 studies where we could not find information about external peer reviews, as they appeared to be published exclusively on the manufacturer's homepage. The greater question of external validity could not be answered with designs that cover only a few months of usage (md: 3m; IQR: 1m) or short follow-ups (md: 6m; IQR: 6m). Only evaluations with a longer-term application under everyday conditions in post-marketing studies could help to address this aspect properly.

5. Conclusions

All 15 applications permanently listed in the German DiGA repository provided evidence supported by RCTs. However, on closer inspection, some of the underlying studies have at least debatable limitations regarding their validity. The quality of the studies already available, nevertheless, exceeds the low minimum requirements for inclusion in the repository under federal law. Nonetheless, further studies are needed to improve the strength of evidence on the benefits of apps on prescription. Evidence gaps need to be closed. This is not only true for Germany, but for other countries that have either already implemented or are in the process of designing policies to be applied in the context of health apps [22].

References

[1] Deutscher Bundestag. Digitale-Versorgung-Gesetz – DVG. Fassung vom 08.11.2019, verabsch. im Bundesrat am 29.11.2019; 2019. [cited 2022 Dec 9]. Available from: https://www.bundesrat.de/SharedDocs/drucksachen/2019/0501-0600/557-19.pdf.

[2] Bundesamt fur Justiz. Digitale Gesundheitsanwendungen-Verordnung – DiGAV; 2020. [cited 2023 Jan 1]. Available from: https://www.gesetze-iminternet.de/digav/BJNR076800020.html.
[3] BfArM. DiGA-Verzeichnis; 2022. [cited 2022 Dec 9]. Available from: https://diga.bfarm.de/de/verzeichnis.
[4] NICE. Evidence Standards Framework for Digital Health Technologies. NHS England; 2019. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf.
[5] Steiner S, Lauterbach KW. Development of evidence-based clinical practice guidelines; a model project integrating external evidence and clinical expertise. Med Klin (Munich). 1999 Nov;94(11):643-7.
[6] Berkman ND, Lohr KN, Ansari M, McDonagh M, Balk E, Whitlock E, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US), https://www.ncbi.nlm.nih.gov/pubmed/24404627 (2013).
[7] Klein JP, Berger T, Schroder J, Späth C, Meyer B, Caspar F, et al. Effects of a psychological Internet intervention in the treatment of mild to moderate depressive symptoms: Results of the EVIDENT study, a randomized controlled trial. Psychother Psychosom. 2016 May;85(4):218-28.
[8] Meyer B, Bierbrodt J, Schroder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interventions. 2015;2(1):48-59.
[9] Twomey C, O'Reilly G, Bultmann O, Meyer B. Effectiveness of a tailored, integrative Internet intervention (deprexis) for depression: Updated meta-analysis. PLoS One. 2020 Jan;15(1):e0228100.
[10] Pöttgen J, Moss-Morris R, Wendebourg JM, Feddersen L, Lau S, Köpke S, et al. Randomised controlled trial of a self-guided online fatigue intervention in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2018 Sep;89(9):970-6.
[11] Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care. 2015 May;38(5):776-83.
[12] Ebenfeld L, Lehr D, Ebert DD, Kleine Stegemann S, Riper H, Funk B, et al. Evaluating a Hybrid Web-Based Training Program for Panic Disorder and Agoraphobia: Randomized Controlled Trial. J Med Internet Res. 2021 Mar;23(3):e20829.
[13] Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: A randomized controlled trial. J Med Internet Res. 2016 Jan;18(1):e21.
[14] Zarski AC, Berking M, Ebert DD. Efficacy of internet-based treatment for genito-pelvic pain/penetration disorder: Results of a randomized controlled trial. J Consult Clin Psychol. 2021 Nov;89(11):909-24.
[15] Kalmeda. Tinnitus-Studie; 2022. [cited 2022 Dec 10]. Available from: https://www.kalmeda.de/fachkreise/tinnitus-studie.
[16] Kramer R, Köhne-Volland L, Schumacher A, Köhler S. Efficacy of a Web-Based Intervention for Depressive Disorders: Three-Arm Randomized Controlled Trial Comparing Guided and Unguided Self-Help With Waitlist Control. JMIR Form Res. 2022 Apr;6(4):e34330.
[17] Lorenz N, Heim E, Roetger A, Birrer E, Maercker A. Randomized controlled trial to test the efficacy of an unguided online intervention with automated feedback for the treatment of insomnia. Behav Cogn Psychother. 2019 May;47(3):287-302.
[18] Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided Internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med. 2017 Jan;47(1):67-80.
[19] Weise H, Zenner B, Schmiedchen B, Benning L, Bulitta M, Schmitz D, et al. The Effect of an AppBased Home Exercise Program on Self-reported Pain Intensity in Unspecific and Degenerative Back Pain: Pragmatic Open-label Randomized Controlled Trial. J Med Internet Res. 2022 Oct;24(10):e41899.
[20] Zill JM, Christalle E, Meyer B, Harter M, Dirmaier J. The effectiveness of an internet intervention aimed¨ at reducing alcohol consumption in adults. Dtsch Arztebl Int. 2019 Feb;116(8):127-33.
[21] Forkmann K, Roth L, Mehl N. Introducing zanadio – A Digitalized, Multimodal Program to Treat Obesity. Nutrients. 2022 Aug;14(15).
[22] Essén A, Stern AD, Haase CB, Car J, Greaves F et al. Health app policy: international comparison of nine countries' approaches. *NPJ Digit Med* 2022; 5: 31.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230166

The Inverse Data Law: Market Imperatives, Data, and Quality in AI Supported Care

Saleem AMEEN[a,1], Ming Chao WONG[b], Kwang Chien YEE[a], Christian NØHR[c] and Paul TURNER[b]
[a] *Tasmanian School of Medicine, University of Tasmania, Australia*
[b] *School of Information and Communication Technology, University of Tasmania, Australia*
[c] *Department of Planning, Aalborg University, Denmark*

Abstract. Over the last decade, the explosion of "Big Data" and its fusion with AI has led many to believe that the development and integration of AI systems in healthcare will usher in a transformative revolution that democratises access to high quality healthcare and collectively improve patient outcomes. However, the nature of market forces in the evolving data economy, has started to show evidence that the opposite is more likely to be true. This paper argues that there is a poorly understood "Inverse Data Law" that will exacerbate the widening health divide between affluent and marginalised communities because: (1) data used to train AI systems favour individuals that are already engaged with healthcare, who have the lowest burden of disease, but the highest purchasing power; and (2) data used to drive market decisions around investment in AI health technology favours tools that increase the commodification of healthcare through over-testing, over-diagnosis, and the acute and episodic management of disease, over tools that support the patient to prevent disease. This dangerous combination is more likely to cripple efforts towards preventative medicine, as data collection and utilisation tends to be inversely proportional to the needs of the patients served – the inverse data law. The paper concludes by introducing important methodological considerations in the design and evaluation of AI systems to promote systems improvement for marginalised users.

Keywords. Big data, artificial intelligence, health equity, digital divide, data bias

1. Introduction

In 1971, Julian Tudor Hart, a British General Practitioner, famously proclaimed the inverse care law – the notion that "the availability of good medical care tends to vary inversely with the need for it in the population served" [1]. Tudor Hart critically identified that when medical care was exposed to market forces, inequity ensued, as those who *needed* healthcare the most (the disadvantaged), tended to receive it the least [1,2]. It was purported that the confluence of artificial intelligence (AI), machine learning (ML), Internet of Things (IoT), patient-generated health data (PGHD), and patient-centred eHealth and mHealth applications would signal the end of the inverse care law, by democratising access to high quality, cost-effective healthcare and supporting the goal of universal health coverage (UHC) [3,4]. However, while some individuals have benefited from the digitisation of healthcare, it has become increasingly clear that there

[1] Corresponding Author: Saleem Ameen, E-mail: saleem.ameen@utas.edu.au

exists a digital divide – those who are less likely to successfully engage with healthcare online (e.g., elderly, disabled, those with poorer education, and the culturally and linguistically diverse (CALD) groups), tend to experience the greatest burden of disease [5]. Some have suggested that the evolving "digital-first" model to global health systems has given rise to a "digital inverse care law"; the evidence of which is best captured by the management of the COVID-19 pandemic response, which highlighted how those who lacked the capacity for digital engagement, were more likely to have more severe symptoms, several comorbidities, less support, and a higher mortality rate [6,7].

While many researchers have started to examine the socio-technical, ethical, and moral questions that arise from the nature of interaction with digital health systems [8]; this paper identifies a lesser-known pervasive *"inverse data law"* that emerges as a by-product of the digital inverse care law, which is likely to significantly exacerbate the health divide between the affluent and marginalised communities. The inverse data law argues that in the presence of market forces, data collection and data utilisation in AI-supported care, tends to be inversely proportional to the needs of the patients served. The paper suggests that one approach to minimise the impact of the inverse data law, is to develop new methodologies around the design and evaluation of AI-based health technologies that are more robust at catering to the needs of diverse marginalised users.

2. Data Collection and Representational Biases in AI Health Technologies

ML algorithms are perceived to be immensely powerful because they are predicated on a fundamental assumption that embedded within large data distributions exists a set of computationally acquirable relationships that can disentangle the factors of variation that map X inputs to Y outputs, when framed as an optimisation problem. While this approach to computation has stimulated several healthcare innovations across screening, diagnostic, and therapeutic pathways [9,10], it has also raised serious concerns around its potential to amplify pre-existing biases within datasets, given that the main objective of the algorithms are to maximise signals that reinforce their distributions [11–13].

Bias manifests in elusive ways and is not a feature that can be eliminated by "Big Data", as it is highly influenced by the interrelationships between the nature of representation and engagement in existing healthcare services, and by the broader socio-organisational constructs that contribute to the widening of health disparities when the complexity of an individual is reduced into a function of a group for data classification purposes (e.g., 'Hispanic', 'Asian', or 'Black') [14]. It is for this reason that researchers that audit large datasets and/or their AI implementations, tend to consistently return the same disappointing results – underdiagnosis or misdiagnosis along the lines of race/ethnicity, sex, age, and insurance type, as was observed in a recent audit of the largest publicly available radiology datasets for chest X-ray predictions [15].

In certain disease contexts, these representational biases are implicit to the nature of the disease distribution itself. For example, even though rare diseases affect more than 300 million individuals globally [16], each condition is characterised as low prevalence, affecting fewer than one in 2,000 people. Often, these diseases manifest with complex heterogeneity, overlapping phenotypes, numerous clinical subtypes, and unknown molecular mechanisms [17]. This characterisation is the antithesis of the big data context that medical AI success traditionally thrives on [18]. There is a concern that an over-reliance on such systems over time may alter clinical interaction and shift more human

resources towards episodic care, making underrepresented users with complex needs who demand continuous care ever-more marginalised.

3. AI Supported Care When Data Utilisation is Influenced by Market Forces

Recently, there has been an explosion in companies that offer digital health services across the continuum of prevention, detection, and management of disease that promises to improve health outcomes. However, in a concerning analysis by Cohen et. al. [19], it was revealed that companies in the United States of America (U.S.) that prioritised the acute and episodic management of disease, received significantly more investment than those that were singularly concerned with the prevention of disease. This is despite the fact that preventable chronic health conditions are responsible for at least two-thirds of health-related deaths in the U.S., and disproportionately affects those from disadvantaged backgrounds [20]. These trends suggest that market forces that dictate the dissemination of health technology are typically uninterested in health equity and the needs of the disadvantaged and/or vulnerable populations, as decisions that drive investment revolve around the maximisation of profit [21].

This raises important considerations when the downstream effects of the inverse data law are examined in the context of marginalised citizens receiving AI-supported care that is developed in the presence of market forces. Marginalised individuals participate the least with healthcare (inverse care law), participate even less with digital health services (digital inverse care law), are the least represented by the datasets used in AI development (inverse data law), are the most susceptible to chronic diseases, and are the most disadvantaged when it comes to accessing adequate care. When the key driver of emerging health technology development is consumerism, important ethical questions around how commercial vendors intend to meet the needs of marginalised citizens arise, particularly as promoting behaviours that prevent the need for healthcare is less likely to be endorsed, if there is little alignment with profitability. One potentiality is that vendors may advocate for prevention through over-testing, irrespective of the clinical evidence around its use [22]. It is unsurprising then, that Google's longitudinal "Project Baseline Health Study" (PBHS), aims to monitor maximal longitudinal health data through frequent testing. While its stated objective is to advance biomedical knowledge discovery and "open science" [23], how this fits into Google's commercial health interests and how it impacts health delivery is yet to be seen.

4. Methodological Considerations to Improve AI Health Technologies

The purpose of the Inverse Data Law is to provide a conceptual lens to view potential trends that may arise when developing emerging AI-based healthcare technologies in a market driven system. This section introduces a preliminary discussion around how design and evaluation methodologies may be developed for an era of AI-supported healthcare that aims to benefit marginalised users, while limiting its potential for harm.

Through the product design and development phase, incorporating user-centred design (UCD) principles to capture the needs of diverse user groups could result in systems that better address the concerns of marginalised citizens. However, there remains questions around how to best approach UCD in an era of agile development and disruptive innovation. For instance, given that a consumer's expectation of technology

tends to be constrained by the imaginative boundary of their familiarity with *existing* technologies, is it better for an innovator to engage participants as participatory co-designers for the purpose of ideation, or should the designer attempt disruptive innovation independently and engage participants *after* the design process to evaluate patient acceptability with an iterative view towards design? Furthermore, determining the best way to stratify participants into groups during UCD can be a challenge, particularly as generic group-based definitions tend to fail to capture the heterogeneity of individuals. For example, if one aimed to develop an AI-based intervention to improve colorectal cancer screening amongst under-screened CALD participants, it is likely that demographic differences would yield different technological demands, such that one design choice may benefit one individual at the expense of another within the same group.

Through the evaluation phase, one approach to improve the safety of AI systems, could be to have an independent authority audit and certify AI-based health systems, prior to and during system use, establishing operational constraints on who can consume the technology, based on system evaluation performances. This requires researchers and clinicians to agree upon and develop an evaluation standard for AI health technologies that appropriately considers the nuances of AI system development across: (1) data distributions used through system development, (2) the nature of the technical architecture of the models and their possible constraints, (3) the socio-technical considerations around their use, and (4) the clinical utility of the systems in practice. A possible starting point for such a framework, could be the recently published medical algorithmic audit by Liu and colleagues [24]. This framework, however, requires further development to provide explicit evaluation tools around how to approach different data types, datasets, model architectures, and application contexts, which have different ramifications for evaluation that need to be carefully examined. For instance, generative large language models (LLMs) [25] that have a capacity to (a) authoritatively "hallucinate" misinformation, (b) produce different outputs at each time step, and (c) that may not be privy to the nuances of language across CALD groups; are likely to have different evaluation parameters when compared to a discriminative model used as part of an image processing pipeline.

5. Conclusion and Future Work

This paper argues that there is a lesser-known inverse data law; the view that in the presence of market forces, data collection and data utilisation in AI-supported care, tends to be inversely proportional to the needs of the patients served. It shows how biases in datasets used to train AI systems are not features that can simply be eliminated, as bias emerges out of the contextual human interactions that drives data formation. Disadvantaged patients are the least likely to be represented in the data that underpins AI system development. When this is met with market forces that drive a profit-first strategy, there is a risk that: (a) biases are exacerbated, (b) episodic care becomes valued over long-term care, (c) over-testing becomes preferred over preventative care, (d) support for chronic and rare disease sufferers reduces; and the ones that pay the price, are the ones who need care the most. To curtail the likelihood of such an eventuality, the paper introduces methodological considerations across the design, development, and evaluation phases of AI systems, which will form the basis of our ongoing work.

References

[1] Tudor Hart J. THE INVERSE CARE LAW. The Lancet 1971;297:405–12. https://doi.org/10.1016/S0140-6736(71)92410-X.

[2] The Lancet. 50 years of the inverse care law. The Lancet 2021;397:767. https://doi.org/10.1016/S0140-6736(21)00505-5.

[3] The Lancet Global Health. Breaking the inverse care law. Lancet Glob Health 2021;9:e218. https://doi.org/10.1016/S2214-109X(21)00049-8.

[4] Report of the third global survey on eHealth Global Observatory for eHealth Global diffusion of eHealth: Making universal health coverage achievable 2016.

[5] Salisbury H, Oxford GP. Helen Salisbury: The inverse care law in the digital age. BMJ 2019;364. https://doi.org/10.1136/BMJ.L308.

[6] Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021;11:e053440. https://doi.org/10.1136/BMJOPEN-2021-053440.

[7] Davies AR, Honeyman M, Gann B. Addressing the Digital Inverse Care Law in the Time of COVID-19: Potential for Digital Technology to Exacerbate or Mitigate Health Inequalities. J Med Internet Res 2021;23(4):E21726 Https://WwwJmirOrg/2021/4/E21726 2021;23:e21726. https://doi.org/10.2196/21726.

[8] Demuro P, Petersen C, Turner P. Health "Big Data" Value, Benefit, and Control: The Patient eHealth Equity Gap. Stud Health Technol Inform 2020;270:1123–7. https://doi.org/10.3233/SHTI200337.

[9] Hee Lee D, Yoon SN. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. International Journal of Environmental Research and Public Health 2021, Vol 18, Page 271 2021;18:271. https://doi.org/10.3390/IJERPH18010271.

[10] Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. Lancet Digit Health 2019;1:e13–4. https://doi.org/10.1016/S2589-7500(19)30002-0.

[11] Stuart EA, Dowdy DW. Rising to the challenge of bias in health care AI. Nature Medicine 2021 27:12 2021;27:2079–81. https://doi.org/10.1038/s41591-021-01577-2.

[12] Trahan LH, Stuebing KK, Fletcher JM, Hiscock M, Weiss B, Bellinger DC, et al. Diagnosing bias in data-driven algorithms for healthcare. Nature Medicine 2020 26:1 2020;26:25–6. https://doi.org/10.1038/s41591-019-0726-6.

[13] Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019;9. https://doi.org/10.7189/JOGH.09.020318.

[14] Kauh TJ, Read JG, Scheitler AJ. The Critical Role of Racial/Ethnic Data Disaggregation for Health Equity. Population Research and Policy Review 2021 40:1 2021;40:1–7. https://doi.org/10.1007/S11113-020-09631-6.

[15] Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature Medicine 2021 27:12 2021;27:2176–82. https://doi.org/10.1038/s41591-021-01595-0.

[16] Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics 2019 28:2 2019;28:165–73. https://doi.org/10.1038/s41431-019-0508-0.

[17] Decherchi S, Pedrini E, Mordenti M, Cavalli A, Sangiorgi L. Opportunities and Challenges for Machine Learning in Rare Diseases. Front Med (Lausanne) 2021;8:1696. https://doi.org/10.3389/FMED.2021.747612/BIBTEX.

[18] Mitani AA, Haneuse S. Small Data Challenges of Studying Rare Diseases. JAMA Netw Open 2020;3:e201965–e201965. https://doi.org/10.1001/JAMANETWORKOPEN.2020.1965.

[19] Cohen AB, Dorsey ER, Mathews SC, et al. A digital health industry cohort across the health continuum. Npj Digital Medicine 2020 3:1 2020;3:1–10. https://doi.org/10.1038/s41746-020-0276-9.

[20] Donoso FS. Chronic disease as risk multiplier for disadvantage. J Med Ethics 2018;44:371–5. https://doi.org/10.1136/MEDETHICS-2017-104321.

[21] Horton AE. Financialisation of Care: Investment and organising in the UK and US 2018.

[22] Chauvin A, Lang E, Bossuyt P. When less is more: mitigating the harms and risks of over-testing through shared decision making. Canadian Journal of Emergency Medicine 2021 24:2 2021;24:126–7. https://doi.org/10.1007/S43678-021-00231-7.

[23] Arges K, Assimes T, et al. The Project Baseline Health Study: a step towards a broader mission to map human health. Npj Digital Medicine 2020 3:1 2020;3:1–10. https://doi.org/10.1038/s41746-020-0290-y.

[24] Liu X, Glocker B, McCradden MM, et al. The medical algorithmic audit. Lancet Digit Health 2022;4:e384–97. https://doi.org/10.1016/S2589-7500(22)00003-6.

[25] Bang Y, Cahyawijaya S, Lee N, Dai W, Su D, Wilie B, et al. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity 2023. https://doi.org/10.48550/arxiv.2302.04023.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230167

Streamlining Tangible 3D Printed and Intangible XR Content Creation and Evaluation: The ENTICE Experience

Panagiotis ANTONIOU[a] Efstathios SIDIROPOULOS[b,1],
Natalia STATHAKAROU[c],Charalampos CHATZIMALLIS[d],
Evangelos CHONDROKOSTAS[e], Dimas SUMUNAR[c], Tobias KARLSSON[c],
Sofia LACHANOUDI[a] , Panagiotis DAVID[b], Konstantinos TAGARAS[b],
Annita VARELLA[a], Alkinoos ATHANASIOU[a], James PICKERING[g],
Andreas TOOULIAS[a], Ion KAROLOS[a], Panagiotis-Marios FILIPPIDIS[d],
Eirini SCHIZA[f], Vassilis VOULGARAKIS[d], Charalambos BRATSAS[e],
Vassilis TSIOUKAS[a] , George TSOULFAS[a] and Panagiotis BAMIDIS[a]
[a]*Aristotle University of Thessaloniki, Greece*
[b]*Anaptixiaki Meletitiki Voriou Ellados, Greece*
[c]*Karolinska Institute, Sweden*
[d]*Open Knowledge Greece, Greece*
[e]*VILABS, Greece*
[f]*University of Cyprus, Cyprus*
[g]*University of Leeds, UK*

Abstract. ENTICE aimed to use co-creative methodologies in order to build a solid creation pipeline for medical experiential content. The project has developed and evaluated immersive learning resources and tools aiming to support well-defined learning objectives using tangible and intangible resources (AR/VR/MR, 3D printing) that are highly sought in the fields of anatomy and surgery. In this paper the preliminary results from the evaluation of the learning resources and tools in 3 countries as well as the lessons learnt are presented towards to the improvement of the medical education process.

Keywords. Co-creation, Immersive Learning, Virtual/Augmented reality

1. Introduction

Medical knowledge is crucial and necessitates a strong theoretical foundation in addition to implicit understanding and experience [1]. Currently there is a gap in educationally sound, effective resources for anatomy teaching. In order to achieve this, the need for suitable educational techniques and approaches is constantly emphasized. Immersive experiential technologies flourish in this setting and help to support medical education. Through their enhanced engagement, virtual patients, chatbots, and Virtual, Augmented, or Mixed Reality (VR/AR/MR) in particular have a significant impact on the affective and educational states of healthcare students [1-4]. Additionally, it has been asserted that

[1] Corresponding Author: Efstathios Sidiropoulos, E-mail: stathsid@gmail.com.

students can successfully visualize abstract rules in concrete ways by taking cues from diverse instructional content [5,7, 8]. The sensory immediacy of these technologies leads to an intuitive anchoring of the key knowledge for the learner and facilitates paradigm building, which in turn enables learners to acquire robust and deep knowledge while minimizing the possibility of establishing or maintaining conceptual errors [4]. 3D printing is the latest tool in the arsenal of healthcare practitioners that visualizes tissue damage and assists in therapeutic diagnosis and intervention through detailed localization of it. This technique makes use of 3D digital models created from medical imaging data, offering pre-surgical training or diagnostic visualization with an unheard-of level of immediateness. Participatory Design (PD) emphasizes the value of balance, receptivity, and non-hierarchical relationships [9, 11] by allowing stakeholders to be fully involved in the design and production process [6,12]. The co-creation process makes it possible to identify [10] knowledge distribution as one of the major factors influencing co-creation efficiency. It has been stated that working in a co-design environment with interdisciplinary actors facilitates the production of more original thoughts and ideas [11].
Through integration of learning objectives in the design and implementation pipeline of digital content creation, ENTICE produces increased quality content, thus improving and thus improve the quality and efficiency of education and training by the use of it. Additionally, the co-creative pipeline, which has been formulated for educational-centric digital resources, intrinsically fosters creativity, innovation and promote entrepreneurship. The aim of this paper is to present the participatory process for the development of tangible/intangible educational resources and the preliminary evaluation results in 3 countries. The aim of this paper is to present the holistic methodology and the first results from an integrative participatory approach for development of immersive VR/AR/MR and 3d print resources.

2. Methods

The core innovation of the project is the validation of a systematic, integrative approach for education-centric experiential teaching episodes and resource content creation for 3D printed models, and AR/VR/MR resources. This approach consists of three main aspects: a) Tangible/Intangible integration in immersive educational resource design, b) Learning Objectives-centered development of educational episodes with immersive resources c) Prototyping Knowledge Engineering methodology for education-centric immersive resources.

The fabrication of 3D printed models suitable for medical educational training began with the segmentation and 3D modelling of anonymized DICOM digital images of real patients in order to optimally depict the human anatomy. This modelling was accomplished using separation techniques of each anatomical detail, with gradation in separate layers, for the representation of which various textures and colors were employed. During the digital reproduction of the final 3D-printed models (Fig.1), specially designed fiducial markers placement regions were generated for the purpose of enhancing their appearance using AR mobile applications. In order to preserve anatomical information, these fields were generated on each surface of each human organ. The student can reveal the fiducial markers by using a specialized holder to remove the corresponding region. Magnets are used to keep the regions on the model enabling easy

attachment and detachment. Magnet locations were also fabricated in other areas of each model, primarily to facilitate the detachment and fusion of diverse anatomy.

Figure 1. 3d printed models

For the development of the XR resources we used Unity3d and combined it with more technologies depending on the tasks at hand. The result was a multi-user exploratory environment, a virtual anatomy installation where anatomically correct models were digitally augmented by labels and panels describing details (Fig.2). To make the conversion to multiuser we used Photon Engine. Photon is built on top of a globally distributed infrastructure, which means it is able to handle a large number of users and maintain a high level of performance even under heavy load. This is especially important for Unity3D projects that require real-time communication and synchronization, such as multiuser applications. In order to handle the AR part of the project we used Vuforia Engine. Vuforia is a very well optimized AR platform that gave us the opportunity to handle multiple targets at the same time and render overlayed 3D objects on top of our tangible resources.

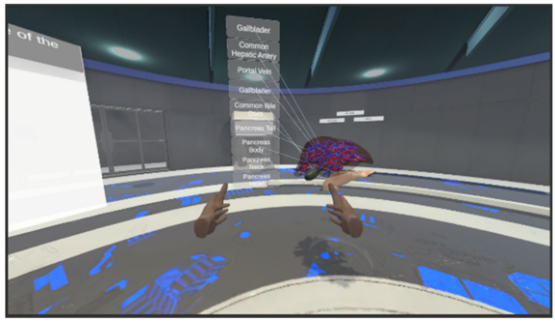


Figure 2. ENTICE VR environment

One of the core innovations of ENTICE was the enrichment of the XR resources using semantic annotations. To accomplish this the first step was to model XR medical resources conceptually and thus a Knowledge Graph (KG) was created by transforming to RDF, medical terms from the UMLS metathesaurus, as well as UX terms for annotating immersive AR/VR/MR content, that were eventually linked and further described using the ENTICE Core Ontology (ECO).

Finally, a qualitative analysis was performed for the initial evaluation of the resources. The qualitative evaluation consists of semi-structured interviews around three main axes: usability, technology acceptance, and perceived usefulness of the resources. Each one of these axes includes more specific topics and sub-topics regarding the questions asked during the interviews. The Unified Theory of Acceptance and Use of Technology (UTAUT) model selected for evaluating the technology acceptance axes [13]. The UTAUT Questionnaire in this evaluation consists of 25 items. Majority of the questions adopted from previously validated research related to VR acceptance. We also aim to evaluate achievements of learning outcomes (LO) of all VR episodes of the study.LO evaluation intends to measure students' knowledge upon completing their VR interaction.

3. Results

While technical and design innovations are well within the methodological purview of this work, the evaluation of the resources are the core results of the holistic approach. The ENTICE pre-pilot was launched to evaluate the virtual reality (VR) resources collecting qualitative data. The developed resources upon first feature complete release were sent for testing in three medical institutions-project partners, in Sweden (Karolinska Institutet), In Cyprus (University of Cyprus) and in Greece (Aristotle University of Thessaloniki). **In Sweden**, eight students participated in a pre-pilot think-aloud session. Preliminary results from the session demonstrated that the visibility of system status should be improved. During the VR interaction, lack of appropriate feedback provoked students pressing similar buttons multiple times. In addition, fostering a sense of freedom and confidence in VR learning are deemed important. Eventually, help and documentation related to VR is necessary. Students felt convenience when having clear instructions on how to interact with the VR. Furthermore, provisioning technical support during the learning episodes is considered valuable by the students. **In Greece**, 4 surgery residence doctors and one general surgeon specialist participated in a pre-pilot facilitated think aloud session. Participants, being already familiar with the material offered feedback on each explained feature as was explained to them. After the experience, semi-structured interviews were conducted to receive even more targeted feedback on the specific evaluation themes. Overall, feedback was positive with all residents expressing the need for curricular integration of these resources, as well as offering expansion ideas for future upgrades of the resources. **In Cyprus**, six medical students (6th grade), one clinical skills teacher and one technician participated in a pre-pilot semi-structured interview. The educational facilitator introduced the participants to the features of the resources as they individually experienced them and asked for feedback regarding educational usefulness and possible changes. Afterwards, the teacher conducted the semi-structured interview to receive more specific feedback of the resource and the overall experience of the students. In general, the feedback was positive and medical students were excited and recommended more features to include in the resource.

4. Discussion and Conclusion

This study confirms the importance of improving the quality and efficiency of learning resources. The work presented here and currently being validated by the ENTICE project pilots is an effort to streamline immersive content creation while maintaining

pedagogical focus [1-3]. This process is still an ongoing process. In that context, it is still limited regarding its efficacy validation as the pre-pilots are currently finishing and wider piloting of the resources is still pending in the coming months. Infrastructural results, however, as discussed, are present. Future Lessons learnt from the process and tools proposed and developed during the project will be the essential enablers for further supporting and proliferating this rapid iterative process.

Acknowledgement

This publication has been conducted in the context of ENTICE, a Knowledge Alliances for higher education Project and Co-funded by the Erasmus + Programme of the European Union (612444-EPP-1-2019-1-CY-EPPKA2-KA).

References

[1] Antoniou PE, Arfaras G, Pandria N, Athanasiou A, Ntakakis G, Babatsikos E, Nigdelis V, Bamidis P. Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: cohort study. JMIR Serious Games. 2020 Sep 2;8(3):e17823.

[2] Antoniou P, Arfaras G, Pandria N, Ntakakis G, Bambatsikos E, Athanasiou A. Real-time affective measurements in medical education, using virtual and mixed reality. InBrain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2020 Oct 2 (pp. 87-95). Cham: Springer International Publishing.

[3] Dolianiti F, Tsoupouroglou I, Antoniou P, Konstantinidis S, Anastasiades S, Bamidis P. Chatbots in healthcare curricula: the case of a conversational virtual patient. InBrain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2 2020 (pp. 137-147). Springer International Publishing.

[4] Kyriakidou MR, Antoniou P, Arfaras G, Bamidis P. The Role of Medical Error and the Emotions it Induces in Learning–A Study Using Virtual Patients. InBrain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2 2020 (pp. 3-12). Springer International Publishing.

[5] McGurk M, Amis AA, Potamianos P, Goodger NM. Rapid prototyping techniques for anatomical modelling in medicine. Annals of the Royal College of Surgeons of England. 1997 May;79(3):169.

[6] Ordanini A, Pasini P. Service co-production and value co-creation: The case for a service-oriented architecture (SOA). European Management Journal. 2008 Oct 1;26(5):289-97.

[7] Pottle J. Virtual reality and the transformation of medical education. Future healthcare journal. 2019 Oct;6(3):181.

[8] Pugliese L, Marconi S, Negrello E, Mauri V, Peri A, Gallo V, Auricchio F, Pietrabissa A. The clinical use of 3D printing in surgery. Updates in surgery. 2018 Sep;70:381-8.

[9] Ranjan KR, Read S. Value co-creation: concept and measurement. Journal of the academy of marketing science. 2016 May;44:290-315.

[10] Shen CW, Ho JT, Ly PT, Kuo TC. Behavioural intentions of using virtual reality in learning: perspectives of acceptance of information technology and learning style. Virtual Reality. 2019 Sep 1;23:313-24.

[11] Trischler J, Dietrich T, Rundle-Thiele S. Co-design: from expert-to user-driven ideas in public service design. Public Management Review. 2019 Nov 2;21(11):1595-619.

[12] Troye SV, Supphellen M. Consumer participation in coproduction:"I made it myself" effects on consumers' sensory perceptions and evaluations of outcome and input product. Journal of marketing. 2012 Mar;76(2):33-46.

[13] Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS quarterly. 2012 Mar 1:157-78.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230168

On the Effective Dissemination and Use of Learning Objectives Catalogs for Health Information Curricula Development

Oliver J. BOTT[a,1], Ursula BERGER[b], Nicole EGBERT[c], Carolin HERRMANN[b], Birgit SCHNEIDER[d], Björn SELLEMANN[c], Cord SPRECKELSEN[d], Brigitte STRAHWALD[e], Julian VARGHESE[f] and Alfred WINTER[d]
[a]*Working Group Curricula of Medical Informatics of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V., Germany*
[b] *Working Group Teaching and Didactics of Biometry of the International Biometric Society - German Region (IBS-DR) and GMDS e.V., Germany*
[c] *Working Group Nursing Informatics of the GMDS e.V., Germany*
[d] *SMITH - Joint Expertise Center for Teaching (SMITH-JET), Germany*
[e] *Working Group Teaching in Epidemiology of the GMDS e.V., Germany*
[f] *Working Group Medical Informatics Education in Medicine of the GMDS e.V., Germany*

Abstract. Catalogs of competency-based learning objectives (CLO) were introduced and promoted as a prerequisite for high-quality, systematic curriculum development. While this is common in medicine, the consistent use of CLO is not yet well established in epidemiology, biometry, medical informatics, biomedical informatics, and nursing informatics especially in Germany. This paper aims to identify underlying obstacles and give recommendations in order to promote the dissemination of CLO for curricular development in health data and information sciences. To determine these obstacles and recommendations a public online expert workshop was organized. This paper summarizes the findings.

Keywords. learning objectives, competencies, catalogs, scientific societies, medical informatics, biometry, epidemiology, medicine, nursing informatics, curricula

1. Introduction

Competencies, in the form of statements of the desired knowledge or skills a student should possess upon completion of a course, or learning objectives, which focus on the broad goals of a course, are common elements of course design. In what follows, we use LO as an abbreviation for competence-based learning objectives. LO Catalogs (CLO) have been introduced and promoted as a prerequisite for high-quality, systematic curriculum development [1]. Since the introduction of the National Competence-Based Catalogue of Learning Goals in Medicine (NKLM), medical study programs in Germany have been following this approach with increasing commitment. There are also national initiatives to promote CLO in the fields of epidemiology, biometrics, medical

[1] Corresponding Author: Oliver J. Bott, E-mail: oliver.bott@hs-hannover.de.

informatics, biomedical informatics, and nursing informatics of the German Society for Medical Informatics, Biometry, and Epidemiology (GMDS) [2]. A prominent international example is the recommendations on Education in Biomedical and Health Informatics of the International Medical Informatics Association (IMIA) [3].

While CLO-based curriculum development is used effectively in medicine [4], the consistent use of CLO for Biomedical and Health Informatics, Biometry and Epidemiology still needs to be improved. Furthermore, using CLO merely as bureaucratic checklists undermines their function as a tool for learner-centered teaching, which has sparked debates in the past [5]. Thus, in using CLO in curriculum development, the promising opportunities are countered by some general obstacles as well as some that are likely to be specific to health data and information sciences.

The aim of this paper is to systematically identify such obstacles and derive recommendations in order to strategically promote the dissemination and use of CLO in the GMDS subject areas with a focus on their use in curricular and module development.

2. Methods

We organized a public online workshop as part of the annual conference of the GMDS 2022 to explore these obstacles and develop recommendations. The workshop consisted of three steps, the presentation of best practices by invited keynote speakers and two working sessions, described in more detail below and are illustrated in figure 1.

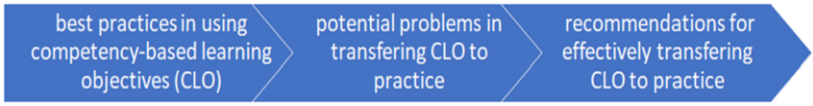


Figure 1. Three steps for bringing CLO to practical use

The workshop started with three keynotes on international experiences with the implementation of CLO in curricula. These lectures focused on the opportunities of recommendation based interprofessional courses in health informatics, the experiences and success factors of implementing the IMIA Recommendations 2010 [6] in a certified master program, and the use of a curriculum information system for transparent, comprehensible and verifiable planning of teaching [7].

The following first working session aimed to identify potential issues in CLO transfer and was conducted using an interactive online whiteboard (Miro) and a videoconferencing tool (Zoom) with the capability of opening breakout rooms, in which the following three topics have been discussed: (1) how do knowledge-oriented and competency-oriented catalogs differ or complement each other, (2) importance and role of scientific societies in the establishment of CLO and (3) individual and structural barriers to using CLO in curricular development with a focus on support and resources.

The second working session began with reports on the status quo and the experiences made so far in implementing their CLO in curricula from five working groups (WG) of the GMDS and the SMITH - Joint Expertise Center for Teaching (SMITH-JET), all of them considered with CLO development and dissemination in Germany: The *WG Medical Informatics Education in Medicine* focuses on competences of medical informatics for medical students, which resulted in a CLO in 2020 [8]. The *WG Teaching and Didactics of Biometry* developed a CLO for teaching biostatistics in medicine in the years 2020-2021 [2]. The *WG Teaching in Epidemiology* published a CLO in 2018 to facilitate planning and implementation of courses in epidemiology [2]. The *WG Nursing*

Informatics published recommendations for core competence fields in nursing informatics in 2017 for use at all levels of educational measures in nursing [2]. The *WG Curricula for Medical Informatics* published a CLO for bachelor programs in (bio-)medical informatics and medical information management in 2021 [2] based on the IMIA Recommendations 2010 [6]. Within the SMITH project of the German Medical Informatics Initiative (MI-I [9]), three master degree programs were developed based on the IMIA Recommendations 2010 [6], the NKLM-MI [8] and the BMHI learning objectives catalog [10] developed by SMITH-JET.

This second working session was carried out technically like the first, but aimed at the development of recommendations for the effective transfer of CLO using the following topics: (1) requirements for the structure, scope and content of CLO with regard to their usability for the development of curricula, (2) recommendations on the role of the scientific societies when establishing CLO and (3) incentives and individual or structural requirements to be created for the use of CLO in curricular development.

3. Results

Identified obstacles and challenges of CLO use in curricular development:

- Differences of knowledge-oriented and competency-oriented catalogs.
  (1) Both approaches complement each other usefully. (2) Competency-based catalogs are more difficult to use for the development of specific training programs if the addressed range of competencies is not clearly defined.
- Importance and role of scientific societies (SCS) in the establishment of CLO.
  (1) SCS often underestimate the human and technical efforts and corresponding costs for the development, maintenance and distribution of CLO.
- Individual and structural barriers to using CLO in curricular development with a focus on support and resources.
  (1) The limited amount of dedicated teaching hours of the medical curricula at many university sites makes it impossible to cover all aspects of a CLO. (2) There is often a lack of training courses, advisory support and suitable tools for the use of a CLO. (3) Learning objectives in existing CLO are often too abstract to be assigned to specific courses or course units. (4) Lecturers and deans of studies often are not convinced of the benefits of CLOs. (5) Regularly changing CLO creates pressure to prematurely change the curricula derived from it.

Identified recommendations for implementing CLO-oriented curriculum development

- Requirements for the structure, scope and content of CLO with regard to their usability for the development of curricula.
  - Teachers perspective: (1) CLO should be consistently divided into levels and LO should be described in a differentiated manner. (2) Competency levels should be clearly separated and competency descriptions should be clearly assigned to a competency level. (3) LO should be formulated in a way that it is possible a) to verify whether they are considered adequately in curricula or module descriptions and b) to measure their achievement. (4) CLO should be accompanied by recommendations for their application. (5) The usability of CLO would benefit from the provision of best practices resp. implementation examples and teaching materials. (6) Curricula

should reference a CLO in a way that it is possible to compare the courses based on a CLO with regard to the spectrum of competencies they address.
  - Students perspective: (1) By entering LO in search fields, it should be possible to find corresponding course offers (also Europe-wide). (2) To achieve this, an automated assignment to the elements of a CLO should be offered when describing a degree program or a course. (3) The results of student evaluations should accompany the development of LO and their implementation in courses and curricula. The evaluation of the implementation should include assessments of the learning success.
  - Employers perspective: (1) Employers should be able to create job advertisements based on CLO. (2) Courses should be searchable for employers using CLO-based profiles (also Europe-wide).
  - Miscellaneous: (1) LO descriptions should be individually accessible and usable (linked open data). (2) Visual representation methods should be offered to ease understanding and implementing a CLO. (3) When developing a CLO, a pragmatic approach should be taken with only a few levels of verb-driven competencies.
- Recommendations on the role of the scientific societies (SCS) when establishing CLO.
  (1) SCS should be aware of their responsibility for providing CLO and actively support their development, maintenance and quality assurance as well as promote their use through suitable, in particular communicative, measures. (2) For this purpose, fixed contact persons should be determined to whom the organization of regular maintenance and support of the CLO is delegated over time. (3) For the development and maintenance of CLO as well as the provision of suitable tools for their maintenance, dissemination and application, the provision of sufficient human, technical, and financial resources by the SCS is necessary. (4) The SCS must decide whether the CLO should be normative, recommendatory, or descriptive in nature. (5) SCS should actively engage in community building to ensure adequate participation of the scientific community during CLO development, e.g. by commenting on CLO drafts. (6) SCS should focus on subject-related training, advanced training and further education in particular to strengthen the importance of the specialist disciplines. To this end, they should take an active part in negotiating the scope of subject-specific teaching, with reference to CLO. (7) SCS should offer or mediate advice on the use of a CLO. (8) The provision of tools for the maintenance and use of CLO should be flanked by SCS by offering training for their application.
- Incentives and individual or structural requirements to be created for the use of CLO in curricular development.
  (1) For the didactically sound implementation of CLO by teachers and program managers, training courses and advisory services should be created. (2) Further training should be focused on the range of available CLO and their application in the discipline concerned. (3) In addition, suitable tools for the CLO-based development of curricula and modules should be provided, that allow to view and search the CLO and map the learning objectives or competencies to modules. Corresponding developments, e.g. the German MI-I [9] should be considered. (4) These tools should also be easy to use and time-efficient. In this respect, they should be regularly evaluated. (5) For efficient and effective use of the tools, appropriate training courses should be provided for teachers and

program managers. (6) Tools such as act-e [7] should primarily be used in the (re-)design of curricula. To avoid rejection of CLO and tool support, their use for monitoring teaching should not be the main focus. (7) It should be invested in measures to increase acceptance that make the effort and benefits of using a CLO and the supporting tools transparent (e.g. by providing best case studies).

4. Discussion

The recommendations developed, while generally not specific to health care, focus on the current development of CLO in this field. That country-specific strategies are particularly needed for effective dissemination of CLO in curriculum development is highlighted in, for example, [3] and [11]. The recommendations developed are based on individual experiences of a limited number of experts, which we gathered in a workshop. In order to gain further insights into the application of CLOs, problems that arise and possible solutions, further empirical research is needed based on techniques like modified RAND or Delphi. This is all the more important as CLO play an increasingly important role - whether as part of legal regulations e.g. in the education of physicians, or in the context of accreditation of educational programs by international scientific societies like EFMI and IMIA. The international perspective is especially important in Europe in order to achieve a Europe-wide comparability of educational programs. For CLO, this means that a simple translation of national CLO is not sufficient, but that semantic references to internationally agreed CLO must be established. This challenge must be met by national and international scientific societies.

References

[1] High Level Group on the Modernisation of Higher Education: report to the European Commission on improving the quality of teaching and learning in Europe's higher education institutions, Publications Office, 2014, https://data.europa.eu/doi/10.2766/42468
[2] GMDS. Lernziel- und Kompetenzkataloge der GMDS [Internet]. GMDS e. V. [cited 2023 Jan 2]. Available from: https://www.gmds.de/aktivitaeten/fachbereichsuebergreifend/lzk-curricula/
[3] Bichel-Findlay J., Koch S., Mantas J, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics: Second Revision, Int. J. Med. Inform., Volume 170, 2023, 104908, ISCSN 1386-5056, https://doi.org/10.1016/j.ijmedinf.2022.104908.
[4] Lammerding-Koeppel M, Fritze O, Giesler M, et al. Benchmarking for research-related competencies – a curricular mapping approach at medical faculties in Germany. Medical Teacher. 2018;40(2):164–73.
[5] Loughlin C, Lygo-Baker S, Lindberg-Sand Å. Reclaiming constructive alignment. European Journal of Higher Education. 3. April 2021;11(2):119–36.
[6] Mantas J, Ammenwerth E, Demiris G, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. Methods Inf Med. 7. January 2010;49(2):105–20.
[7] Teaching & Learning Centre (TLC) of the Amsterdam UMC (location AMC). act-E [Internet]. act-E | Unravel your curriculum. [cited 2022 Dec 3]. Available from: https://www.act-e.nl/
[8] Varghese J, Röhrig R, Dugas M. Which competencies in medical informatics are required by physicians? An update of the catalog of learning objectives for medical students. GMS MedInform Biom Epidemiol. 2020;16(1):Doc02. DOI:10.3205/mibe000205
[9] SMITH [Internet]. Smith.care. [cited 2022 Nov 30]. Available from: https://www.smith.care/
[10] BMHI learning objective catalogue, Version 5. Available from: https://lern-z.imib.rwth-aachen.de/hilona/index.php/BMHI-Version-5
[11] Jaspers MW, Mantas J, Borycki E, Hasman A. IMIA accreditation of biomedical and health informatics education: current state and future directions, Yearbook in Medical Informatics 26 (1) (2017) 252–256. https://doi.org/10.15265/IY-2017-011.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230169

Use of Digital Games for Educational Purposes Among Medical and Paramedical Sciences Students, Mashhad, Iran

Masoumeh SARBAZ[a], Alireza BANAYE YAZDIPOUR[a,b,c], Seyyedeh Fatemeh MOUSAVI BAIGI[a,d], Ali DARROUDI[a,d] and Khalil KIMIAFAR[a,1]
[a]*Department of Health Information Technology, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran*
[b]*Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran*
[c]*Students' Scientific Research Center (SSRC), Tehran University of Medical Sciences, Tehran, Iran*
[d]*Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran*

Abstract. This study aimed to investigate the use of digital games for educational purposes among medical and paramedical sciences students at Mashhad University of Medical Sciences (Northeast of Iran). This cross-sectional study was conducted from July 2018 to January 2019. The research population was all students of the school of medicine and school of paramedical sciences at Mashhad University of Medical Sciences (n = 496). The research tool was a researcher-made questionnaire based on a literature review. The validity of the questionnaire was confirmed by its content validity, and the reliability of the questionnaire was evaluated based on the test-retest method (r = 0.82). In this examination of medical and paramedical sciences students' attitudes and perspectives, some novel preliminary insights into the applications, advantages, disadvantages, and features of digital games in education emerge. Overall, the findings showed that the use of interactive digital games can increase students' motivation for learning and make the learning process more attractive for students. This study was approved by the ethical committee of MUMS (approval number IR.MUMS.REC.1397.151).

Keywords. Digital game, medical education, gamification, e-learning

1. Introduction

A digital educational game for specific purposes of medical education is an electronic game that enables the interaction of professionals and students with a user interface in an offline and/or online mode reference. Digital games are developed for use on iPads, smartphones, computers, tablets, etc. Digital games lead to the strengthening of the attitude and motives of players towards learning through bolding the feeling of competition, using encouragements such as earning scores, and increasing the difficulty

[1] Corresponding Author, Department of Health Information Technology, School of Paramedical Sciences, Mashhad University of Medical Sciences, Azadi Square, Pardis Daneshgah, Mashhad, IR Iran, Tel: +98-5138846728, E-mail: Kimiafarkh@mums.ac.ir

[1]. In addition, the game allows the player to quickly assess their progress with specific goals and immediate and reactive feedback [2]. Although extensive research has been conducted on the application of digital games in medical education, according to the researchers' knowledge, there are few studies that cover the attitude and use of digital games among medical students. Therefore, this study was conducted with the aim of investigating the use of digital games for educational purposes among medical and paramedical students at Mashhad University of Medical Sciences.

2. Methods

This cross-sectional study was conducted from July 2018 to January 2019. The research population was all students of the school of medicine and the school of paramedical sciences at MUMS. Mashhad (in the northeast of Iran) is the second-largest city in Iran after Tehran. MUMS is one of the best medical universities in Iran, operating under the auspices of the Ministry of Health, Treatment, and Medical Education in Mashhad. This research was confirmed by the ethical committee of MUMS (approval number IR.MUMS.REC.1397.151). This research was conducted using a researcher-made questionnaire and a literature review [10, 13-16]. The main components of the questionnaire include the time spent on playing digital games for educational purposes, the impact of using digital games on academic achievement, students' attitudes towards the application of digital games in education, students' attitudes towards the advantages and disadvantages of digital games, and students' attitudes about the features of digital games that were ideal for educational purposes. The validity of the questionnaire was approved by its content validity, and the questionnaire was examined by five faculty members (two medical informatics, two health information management specialists, and a social medicine specialist). The reliability of the questionnaire was evaluated based on the test-retest method ($r = 0.82$). The questionnaire was offered on a voluntary and anonymous basis to 549 medical and paramedical sciences students. The sample size was determined based on the Cochran formula with a 95% confidence interval independently for each community (medical and paramedical sciences schools). A p-value of 0.05 was applied to determine the level of statistical significance. Data were analyzed in SPSS version 11 software using descriptive and analytical statistics.

3. Results

The questionnaire was completed by 496 respondents, with a response rate of 90%. In this study, the majority of participants were female (76.2%) and single (79.8%), and the average age of students was 22.62±2.81 years old. The findings of our study show that among the tools used for digital games for educational purposes, mobile phones had the most use, with 75.7% in medical students and 66.7% among paramedics, and computers in the university had the lowest use, with 84.6% in medical students and game consoles with 73.1% among paramedical sciences students. However, there was no significant difference in the use of digital game tools between medical and paramedical students. The place of use of digital games at home or in the dormitory is the most popular among medical students (69.2%), and paramedical sciences students (60.2%) and the least popular among medical students (54.3% on the way) and paramedical sciences students (55.8%) in the classroom. There was a significant difference between the use of medical

and paramedical sciences students at the beginning of each game (P-value <0.001); most paramedics (44.2%) stated that they use less than ten minutes of digital games for educational purposes. Nevertheless, as they stated, the majority of medical students (36.0%) use digital games for more than ten minutes to an hour for educational purposes. Medical and paramedical sciences students chose digital games as their priorities for learning objectives and specialized educational activities (68.4% and 52.6%), increasing general information (38.9% and 37.8%), and recreation (30.8% and 62.2%). According to Table 1, regarding the effect of digital games on academic achievement, the majority of medical students (51.4%) and paramedics (45.4%) stated that the use of digital games leads to their academic achievement. The highest percentage of agreement with the use of digital games among students was about their use for educational purposes, so 58.7% of medical students (*strongly agree* and *agree*) and 59.8% of paramedical sciences students (*strongly agree* and *agree*) tended to use digital games for educational purposes. There was no significant relationship between gender and willingness to use digital games for educational purposes (P-value = 0.221). The majority of agreements regarding the benefits of digital games were about communicating with other game players while playing and helping to facilitate learning. 93.6% of medical students and 98.4% of paramedical sciences students agreed and strongly agreed with communicating with other players during the game. 60.8% of medical students (*strongly agree* and *agree*) and 54.2% of paramedical sciences students (*strongly agree* and *agree*) stated that the use of digital games can facilitate learning. Furthermore, 71.7% of medical students and 75.1% of paramedical sciences students agreed and strongly agreed that it should be different ways to earn points in the game. 63.2% of medical students and 69.1% of paramedical sciences students agreed and strongly agreed that there should be possible to play digital games with different tools.

Table 1. The most important students' attitudes about the application in education, advantages and disadvantages, and characteristics of educational digital games

The most important students' attitudes toward the use of digital games in education								
Questions	**Schools**	**Strongly agree N (%)**	**Agree N (%)**	**Somewhat agree N (%)**	**Disagree N (%)**	**Strongly agree N (%)**	**Mean ± SD**	**P-value**
I like to use digital games for educational purposes.	Medicine	75(30.4)	70(28.3)	67(27.1)	28(11.3)	7(2.8)	2.27±1.10	0.025
	Paramedical Sciences	81(32.5)	68(27.3)	66(26.5)	14(5.6)	20(8.0)	2.29±1.20	
The most important students' attitudes toward the advantages and disadvantages of digital games								
It is possible to communicate with other players during the game.	Medicine	138(55.9)	93(37.7)	9(3.6)	2(0.8)	5(2.0)	1.55± 0.78	0.033
	Paramedical Sciences	164(65.9)	81(32.5)	2(0.8)	1(0.4)	1(0.4)	1.36±0.56	
The use of digital games facilitates learning.	Medicine	74(30.0)	76(30.8)	70(28.3)	23(9.3)	4(1.6)	2.21±1.02	0.002
	Paramedical Sciences	43(17.3)	92(36.9)	81(32.5)	19(7.6)	14(5.6)	2.47±1.04	
The most important students' attitudes about the characteristics of educational digital games								
Ways to earn points should be optional for the individual.	Medicine	71(28.7)	107(43.3)	47(19.0)	19(7.7)	3(1.2)	2.09±0.94	0.027
	Paramedical Sciences	60(24.1)	100(40.2)	59(23.7)	15(6.0)	15(6.0)	2.29±1.08	
	Paramedical Sciences	78(31.3)	74(29.7)	68(27.3)	14(5.6)	15(6.0)	2.25±1.13	
	Medicine	83(33.6)	73(29.6)	50(20.2)	39(15.8)	2(0.8)	2.20±1.09	<0.001

It allows play and continue digital games with different types of tools (such as mobile phones, laptops, tablets, etc.).	Paramedical Sciences	94(37.8)	78(31.3)	54(21.7)	8(3.2)	15(6.0)	2.08±1.12

4. Discussion

4.1. Principal Findings

According to the findings of the present research, the use of these digital educational games among students is mostly for teaching and learning purposes. In this study, most medical and paramedical sciences students stated that digital educational games could have a positive effect on their academic achievement. According to Noraddin et al. [3], most students stated that they tend to use digital educational games for learning purposes because they facilitate the learning of relevant courses. In line with the results of this study, Alastair et al. found that complementary medicine students have a high willingness and ability to use digital technologies as a learning stimulus [4]. We also found that among the tools used for digital games, students preferred to use mobile phones. Mobile-based educational games allow players to play whenever they want. Therefore, when the player has free time, s/he can use it optimally [5, 6]. The results of a comparative survey of the use of mobile-based games compared to other digital gaming tools showed that the majority (54%) of students prefer to play mobile-based games [7]. As shown in the findings of this study, most paramedical sciences students (44.2%) use less than ten minutes of digital games for educational purposes. In contrast, the majority of medical students (36.0%) stated that they use digital games for more than ten minutes of an hour for educational purposes. However, in general, the results of this study showed that students tend to play at different times with a duration of less than one hour. Sitzmann, in his meta-analysis study, found that, based on media comparisons, unlimited access to the game had much better learning outcomes than limited access to the game [8]. Wouters et al. focused on a combination of distance learning and one-time learning. As they explained, this result is plausible because, compared to traditional teaching methods, the effect of digital games on learning is only compensated for after a few practice sessions when the players become accustomed to the game [9]. We found that the majority of students considered the use of digital games as a means of academic achievement and recognized digital games as a tool to facilitate learning. As Fuster-Guilló et al. found, the majority of students reported a positive perception and attitude toward learning through digital games (56% believed that digital games enhanced their learning, and 48% stated that digital games motivated them to learn) [10]. Also in this study, 93.6% of medical students and 98.4% of paramedical sciences students expressed that they agreed and strongly agreed with communicating with other players during the game. Consistent with the results of our study, Kron et al. found that if digital games were multiplayer (97%) and helped to develop patient interaction skills (90%), they were more likely to use multiplayer and interactive digital games (simulators) [11]. Therefore, in the design of digital educational games, student participation increases probably the sense of competition, attraction, and thus better learning.

4.2. Strengths and limitations

One of the strengths of our study was the high response rate of the participants to this survey (response rate of 90%); due to the careful follow-up for the correct response, it had a minimum of missed data. Moreover, we designed a questionnaire that can identify the maximum perceptions and views of students as the main users of digital games for optimal design and policy-making according to their needs. However, our study also had some potential limitations. First, this study Participants self-selected, and as such, this may have contributed to the selection bias. Second, this was a single-center study, which limits the generalizability of the results. Third, most of the participants in this study were young medical and paramedical science students. It is unclear how sufficient this kind of survey would be in an older population such as senior doctors. Despite these limitations, the results from this study provide valuable insights into medical and paramedical science students' perspectives and experiences regarding digital games. Thus, other students' perceptions of digital games in other institutions and universities require further investigation and comparison with these findings.

5. Conclusions

The findings showed that the use of digital games can increase students' motivation for learning and make the learning process more attractive for students.

References

[1] Mousavi Baigi SF, Sarbaz M, Marouzi P, Kimiafar K. Evaluating the impact of digital game on learning medical terminology of paramedical students: Protocol for a randomized controlled trial. InAdvances in Inform, Manag & Technol in Healthcare 2022 (pp. 51-54). IOS Press.
[2] Mousavi Baigi SF, Aval RN, Sarbaz M, Kimiafar K. Evaluation Tools for Digital Educational Games: A Systematic Review. Acta Medica Iranica. 2022 Aug 1;60(8).
[3] Noraddin EM, Kian NT. Academics' Attitudes toward Using Digital Games for Learning & Teaching in Malaysia. Malaysian Online J of Educ Technol. 2014;2(4):1-21.
[4] Gray AC, Steel A, Adams J. Complementary medicine students' perceptions, perspectives and experiences of learning technologies. A survey conducted in the US and Australia. Eur J Integr Med. 2021;42:101304.
[5] Karabatzaki Z, Stathopoulou A, Kokkalia G, Dimitriou E, Loukeri PI, Economou A, et al. Mobile Application Tools for Students in Secondary Education. An Evaluation Study. Inter J of Interact Mob Technol. 2018;12(2).
[6] Alzatma AA, Khader K. Using Mobile Apps to Improve English Speaking Skills of EFL Students at the Islamic University of Gaza. Linguistics. 2020.
[7] Chung L, Prado Leite JCSd. On non-functional requirements in software engineering. Conceptual modeling: Foundations and applications: Springer; 2009. p. 363-79.
[8] Sitzmann T. A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Pers Psychol. 2011;64(2):489-528.
[9] Wouters P, Van Nimwegen C, Van Oostendorp H, Van Der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol. 2013;105(2):249.
[10] Fuster-Guilló A, Pertegal-Felices ML, Jimeno-Morenilla A, Azorín-López J, Rico-Soliveres ML, Restrepo-Calle F. Evaluating impact on motivation and academic performance of a game-based learning experience using Kahoot. Front Psychol. 2019;10:2843.
[11] Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ. 2010;10(1):1-11.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230170

Perceptions of Learning Activities in Electronic Health Record Transition

Thomas SCHMIDT[a,1] and Christian NØHR[b]
[a] *Center for Health Informatics and Technology, University of Southern Denmark, Denmark*
[b] *Danish Center for Health Informatics, Department of Planning, Aalborg University, Denmark*

Abstract. Learning activities are at the front-line of first impressions. In this paper, the education and training program for a large electronic health record transition project is presented. Management, and staff were interviewed before, during, and after implementation on their perception, reception, and benefit of various learning activities. Daily clinical work and obligations complicate adherence to learning programs, and the clinical professions differ in their approach to mandatory activities. Local learning activities empower staff, and planners should consider embedding room for adjustment of learning program during implementation.

Keywords. Electronic Health Records, EHR transition, Education, Training

1. Introduction

A hospital's Electronic Health Record (EHR) system plays an essential role in ensuring timely, proper, and safe care of patients. As such the EHR becomes intertwined with clinical work and ensuring that all staff has a shared understanding of functionality and usage is necessary for any well-functioning hospital. The digitalization of clinical information and knowledge in Scandinavian countries has been ongoing for decades, and thus most EHR implementations can be considered as system transition projects [1]. However, assuming that transition projects are simpler than an implementation project, or are less risky, or complicated, is naïve [2]. EHR system transitions face several challenges and obstacles as the replacement system simultaneously needs to fit into existing perceptions and procedures, while still seeking to expand and extend scope and functionality. Hospitals are expected to maintain continuous service delivery, and daily interdependencies between departments complicates parallel operation of the old and new EHR systems. To smooth this transition as much as possible, education and training of personnel is of crucial importance [3]. These preparations are costly and needs to be well organized and timed, as training and educational activities often stage expectations and attitudes through first-impression exposure to new circumstances of doing work [4]. Planning and executing training of thousands of employees is no small task, and each effort needs to be weighed carefully as the impact of either doing too little, or too much, training in the new systems quickly accumulates. Differentiating between education and

[1] Corresponding Author, Thomas Schmidt, Center for Health Informatics and Technology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark; E-mail: schmidt@mmmi.sdu.dk.

training, where the former is the process of learning and acquiring information, and the latter is the process of teaching specific skills may serve as a helpful distinction when assessing the outcome of skill acquisition initiatives. Samadbeik et al. [5] identifies five types of training methods: one-on-one training, peer-coach training, classroom training, computer-based training, and blended training. As well as multiple strategies and techniques for training: cased-based, process-based, role-based, feedback and support, functionality, flipped-classroom, and team-based learning. Of these options, classroom-based training is the most widely used format, and case-based training the most common technique.

In this paper, an initial analysis of how staff in the Region of Southern Denmark (RSD) perceived the training and educational program planned for their transition to a new EHR system is presented. The feedback is assessed using the terminology of [5], with the intention of inspiring future planning of learning programs. The paper also frames the current approach in the light of a larger, cross-regional research project where further investigations into broader and long-term effects of training and educational activities for EHR transition are pending.

2. Background

Both RSD and the North Denmark Region (NDR) implemented the same EHR system in 2022, replacing two different existing systems which had been in use for more than a decade. RSD serves a population of 1.4M citizens with a secondary healthcare sector organized into four hospital units, as well as one psychiatric hospital distributed amongst the other units. Figure 1 shows the organization and responsibilities of the learning program. All staff with EHR access were expected to complete a mandatory e-learning educational program prior to the transition date. This training consisted of interactive learning modules and intended to ensure a uniform introduction to the basic functions of the new EHR system. In addition to the e-learning activities; secretaries, and staff involved with use of speech recognition (SR), as well as specially appointed employees known as resource-staff, were assigned to class-based teaching. These employees were tasked with acting as ambassadors and were committed to assisting their colleagues during the transition.

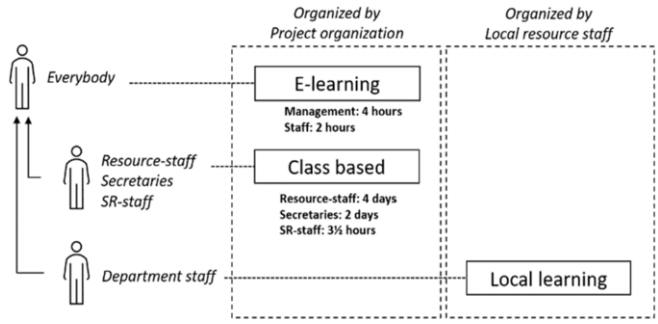


Figure 1. Overview of the various learning activities

Spanning four days, class-based teaching was significantly more extensive and sought to instill knowledge of the new EHR system and provide foundational insight into

the new EHR. The class-based activities were organized as short-talks and assignments to train the participants in core system functionality such as booking patients, managing referrals, handling medications, and dealing with communication. Both e-learning and class-based activities were organized centrally by RSD. Finally, each department had the option of organizing local learning. Planned and conducted by the departments own resource-staff who utilized their new system expertise in devising case-based teaching. Depth and span of local learning activities thus varied by department.

3. Methods

The implementation projects in both regions were studied by an independent research group aiming to investigate the influence of leadership on a large scale EHR transition project – see [6] for a description of the overall project aim. The project consisted of three main activities; 1) regional surveys based on change management and technology acceptance models, 2) registry studies on productivity and financial impact, and 3) a qualitative case study seeking to capture expectations and experiences with the EHR systems amongst clinicians before, during, and after implementation. The case study includes time-motion observations of clinical work amongst physicians, nurses, and secretaries, as well as interviews with staff from a medical, a surgical, and a psychiatric department. 18 clinicians, 7 department managers, and 4 top managers were included in semi-structured interviews in sessions varying from 15-60 minutes. The interviews with clinicians and department managers were analyzed to uncover perception and impression of the training and educational program. The interviews were compared with preliminary results from a survey to contrast the comments from the informants. The survey was distributed by the research group, and the data included in this paper had a response rate of 27% (n=2910) before implementation, and 21.8% (n=2290) after. Only responses regarding the expectations and perceived value of learning activities are included.

4. Findings

Interviews were coded according to themes on perceived quality and effectiveness of the various learning activities, the extent to which it prepared the participants for the new system, and how they coped with issues emerging during the implementation process. One of the key obstacles in enrolling thousands of busy employees in any mandatory activity is the likelihood that more pressing clinical work takes precedence. This issue was especially emphasized by the department managers:
"Just today one of my surgeons informed me that they had to postpone their e-learning due to the scheduling of an acute surgery. That is just the reality of our doctors – even though we plan educational activities – there is always something else they would rather be doing. Which often results in the e-learning being skipped altogether." (Department manager A), and: *"The e-learning is so extensive – 4 hours is a long time. I myself divided it into five chunks, simply because it drained me. In total I probably spent six hours on it.*" (Department manager B). Several of the secretaries expressed a need to know how the new system affected the other professions, to be ready to assist if problems emerged:

"I've completed all my e-learning modules, and actually also modules for the physicians. Even though I do not deal with medication lists etc., it is still nice for me to know what the physicians need to do to handle them" (Secretary).

Although several concerns that the mandatory activities were too generic were observed, most informants also expressed the benefits of participating in e-learning activities:*" Of course you need hands-on experience. That is how I best learn, and it helps me to be less frightened next Monday when the new system is ready. Instead, I'll be like "oh yes, I recognize this"* (Nurse). Still, hands-on training was highlighted as important to feel assured that work could be conducted as required: *"After completing e-learning, you have a pretty good grasp of the system. We do not know how to work with it, but we know some of the fundamentals. It prepares us for the new system by knowing how it looks and feels. Even so, we still don't know how to use the system to do our work tasks, that is something we lack."* (Secretary).

Due to the sequential roll-out of the new EHR, hospital units following the initial implementation site were able to adjust the learning program. While the mandatory e-learning program remained unchanged, it quickly became apparent that local learning initiatives should be emphasized by defining department specific use-cases for the local staff to try out: *"During the classroom lectures we were told it was a good idea to plan local learning at our own departments. We have done it our own way; others have probably done it completely different. The physicians constructed their own local learning activities, and the secretaries and nurse resource-staff have built shared content.".* (Resource-staff at 2. Implementation site).

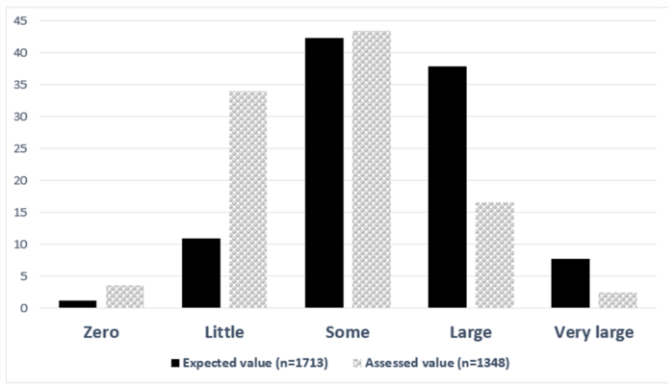


Figure 2. The staff's evaluation of expected and perceived benefits from participating in training programs in percentage of n.

Comparing the interview data with survey findings, both supportive and contrasting indications is found. Prior to participating in the learning activities, the staff was asked about their expectations for their benefit from teaching in EHR in RSD using a five-point scale from zero to very large. Right after the implementation they were asked to rate their benefit from the EHR teaching activities on the same scale. Figure 2 show the results, and it is seen that the training activities did not live up to the expectations as roughly 35% of respondents had high expectations of outcome from participation. In hindsight, the respondents adjusted their self-reported need for training and education and only 19% assessed a high outcome of their training and education. 87% of all respondents participated in the learning activities, reflecting the fact that completion was mandatory and monitored.

5. Discussion

Learning activities is often the point of first contact, and thus also establishes first impressions. As such, planners should pay attention to expectation management and assess the necessity for running follow-up activities. This is best made possible by ensuring ample slack for tweaking activities as well as structure of the learning program to also support targeted education [7]. Here, this was made possible as RSD decided on a gradual roll-out focusing on the somatic hospital units grouped into three stages. Doing so enabled the following implementation units to adjust their learning activities. From our interviews, it appears that staff in general are more satisfied with hands-on training activities. The flexibility of e-learning is a double-edged sword; simultaneously offering flexibility, but at a cost being deferred by more urgent tasks. Furthermore, the mandatory e-learning activities were good at establishing basic familiarity with the system, but the format and self-paced nature, made it difficult to embed into busy schedules. As the same e-learning program was utilized across distinctly different specialties, a lot of informants – especially from smaller departments – expressed that the material was too generic and not adjusted to their circumstances. Local learning initiatives alleviated these concerns.

6. Conclusion

Differences in attitude and coping emerges across professions. E-learning, despite being mandatory, is more easily dismissed and displaced by daily operations. Still, e-learning has offers benefits in terms of establishing familiarity and may be best thought of as a catalyst for later learning activities such as case-based scenarios and peer-coach training. Although departmentalized local learning initiatives are well received, and empowers local change agents, management should consider ways to formalize a foundational set of cases and still encourage local customization, now at a lower cost.

References

[1] Schmidt T, Nøhr C, Vingtoft S, Turner P. Next generation EHRs-what problems are these systems aiming to solve?. InITCH 2019 Jan 1 (pp. 370-374).doi:10.3233/978-1-61499-951-5-370.

[2] Huang C, Koppel R, McGreevey III JD, Craven CK, Schreiber R. Transitions from one electronic health record to another: challenges, pitfalls, and recommendations. Applied clinical informatics. 2020 Oct;11(05):742-54. doi:10.1055/s-0040-1718535.

[3] McAlearney AS, Robbins J, Kowalczyk N, Chisolm DJ, Song PH. The role of cognitive and learning theories in supporting successful EHR system implementation training: a qualitative study. Medical Care Research and Review. 2012 Jun;69(3):294-315. doi:10.1177/1077558711436348.

[4] Heponiemi T, Gluschkoff K, Vehko T, Kaihlanen AM, Saranto K, Nissinen S, Nadav J, Kujala S. Electronic health record implementations and insufficient training endanger nurses' well-being: Cross-sectional survey study. Journal of Medical Internet Research. 2021 Dec 23;23(12):e27096. doi:10.2196/27096.

[5] Samadbeik M, Fatehi F, Braunstein M, Barry B, Saremian M, Kalhor F, Edirippulige S. Education and Training on Electronic Medical Records (EMRs) for health care professionals and students: A Scoping Review. International Journal of Medical Informatics. 2020 Oct 1;142:104238. doi:10.1016/j.ijmedinf.2020.104238.

[6] Hansen MB, Kidholm K, Nøhr C, Schmidt T, Elmholdt KT. Model for Evaluating the Implementation of a Third Generation EHR System. InContext Sensitive Health Informatics: Sustainability in Dynamic Ecosystems 2019 (pp. 141-147). IOS Press. doi:10.3233/SHTI190153.

[7] Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health informatics journal. 2020 Mar;26(1):592-612. doi:10.1177/1460458219839613.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230171

Factors to Consider when Introducing Digital Social Activities to Older Persons with Home Care

Madeleine BLUSI[a,1], Helena LINDGREN[b] and Ingeborg NILSSON[c]
[a]*Department of Nursing, Umeå University, Sweden*
[b]*Department of Computing Science, Umeå University, Sweden*
[c]*Department of Community Medicine and Rehabilitation, Umeå University, Sweden*
ORCiD ID: Madeleine Blusi orcid.org/0000-0003-1428-1950, Helena Lindgren orcid.org/0000-0002-8430-4241, Ingeborg Nilsson orcid.org/0000-0002-8265-5769

Abstract. Social isolation and loneliness have become everyday concerns for populations all over the world as these factors are affecting both physical and mental health in a negative way. Feelings of isolation and loneliness are increasingly acknowledged as a health risk among older persons. ICTs have been recognized as effective tools to combat social isolation among older people. The aim of this study was to explore factors of significance when introducing a tablet-based system providing digital social activities for older persons with home care. Participants were 17 persons, age 70 and older, who lived alone and had assistance from home care. This exploratory study used cross-sectional qualitative data analyzed through thematic analysis. Three themes were generated: 1) lacking vocabulary related to the context, 2) intuitive user interface may replace extensive instructions and 3) unwillingness to commit to a pre-defined measure of performance.

Keywords. Social isolation, social activity, older persons, home care, tablet, digital

1. Introduction

Social isolation and loneliness have become everyday concerns for populations all over the world [1]. These factors are affecting both physical and mental health in a negative way [2,3] and are acknowledged as a health risk among older persons [4]. Participation in activities and social contexts generally decrease in correlation with higher age, often a result of physical limitations preventing individuals from leaving home to participate in social life [5]. In many countries home care is the most common form of care provided for older persons [6]. Home care includes a mix of services provided in the individual's home [7,8]. Social support is one of those services, however decreasing resources have contributed to limiting home care workers' capacity to fulfil social needs for their clients [8,9]. Both quantitative and qualitative evidence have been presented to support that Information and communication technologies, ICTs have been recognized as effective tools that may assists older people to combat social isolation [10]. In a previous research project in northern Sweden a model for enabling online participation in individualized

[1] Corresponding Author: Madeleine Blusi, Department of Nursing, Umeå University, Sweden. E-mail: madeleine.blusi@umu,se

meaningful social activities, IMSA, was developed and evaluated [11]. Through an iterative co-creation process, involving persons aged 70+ years and a multi-disciplinary team of researchers from nursing, occupational therapy and computer science, a variety of online-accessible social activities were tested and implemented in a tablet.

At the time when results from the IMSA study were reported, home care service providers in northern Sweden had identified loneliness and social isolation as major health challenges among care recipients. Innovative solutions to meet the need for social interaction were desired. IMSA had been developed in the same geographic region, for persons of similar age, however participants in the IMSA co-creation process had been living independently in their own homes and home care recipients had not been represented. The aim of this study was to explore factors of significance when introducing a tablet-based system providing digital social activities for older persons with home care.

2. Methods

This exploratory study used cross-sectional qualitative data. Data analysis were guided by the six-phase approach to thematic analysis as described by Braun and Clarke [12] applying a mainly inductive data coding at semantic level.

2.1. Setting and Participants

The study took place in a municipality located in northern Sweden. Purposive sampling was used to recruit and consecutively include participants. Inclusion criteria: 70+ years old, home care recipient, living in their own home. Exclusion criteria: cognitive- or language impairment, living in a nursing home or care facility. Home care staff identified 17 care recipients who had expressed feelings of loneliness and/or desire for more social activity. The introduction procedure consisted of three steps, each step was carefully planned to be smooth and safe for the participant, Table 1. All visits were in each participants home.

Table 1. The three steps of the introduction process.

Step	How/what	Purpose
1	Visit by home care staff. Oral information that a researcher will come to visit them and show the social activities they desired to learn more about and try.	Information from a person familiar to the participant. Provide name and photo of the researchers to gain knowledge about who will come and knock on their door, safety precaution because local police had advised older citizens to not open doors to strangers. Information about the system.
2	Visit by researcher. Oral information and demonstration of the demo social activities implemented in the tablet.	Describe and show the tablet and the activities. Information about the system. Becoming familiar with the tablet and the social activities of choice. Hold and touch the tablet. Trying the demo activities (open, participate in and exit the activity). Allowing the participant to explore activities without pressure.
3	Visit by researcher	Independently open and participate in activity of choice, with researcher by their side as support.

2.2. Design and Data Collection

Cross-sectional data was collected during observations of, and dialogues with, participants during introduction of the IMSA-model, described above, with social activities implemented in a tablet with user interface designed to be intuitive. Two types of field notes were collected: a) during visits notes were written on paper and b) immediately after leaving the participant's home, the researcher audio-recorded a verbal description of the session, including what the participant had said and done, and what the researcher's observations.

2.3. Ethical considerations

All persons invited to participate in the study were given written and oral information according to ethical guidelines for research. Informed consent was obtained from all participants prior to introduction. Ethical approval for the study was granted by the Regional Ethical Review Board in Umeå, Sweden (Dnr 2017/50-31).

3. Results

During analyses three themes were generated. Among the 17 participants were 15 (88%) women and two (12%) men. All participants had a mobile phone kept within reach throughout the visit, 14 (82%) used a touchscreen smartphone, five (30%) had a computer. All participants were in frail health conditions with multiple diseases. Four of them (24%) used a wheelchair for mobility. Three participants (18%) were able to leave their home without assistance, the other 14 (82%) needed help from another person in order to leave their home.

3.1. Theme 1. Lacking vocabulary related to the context.

A challenge encountered during the first steps of the introduction procedure was that participants did not understand the meaning of words related to the context of digital technology or internet. These words were used for giving information and answering curious questions from participants about the system. Despite daily use of smartphone they did not realize that those were connected to the internet. Neither did they understand terms like wi-fi, broadband, tablet, router, connection, touchscreen etc. Provision of "correct" information was mainly left out and focus was on practical demonstrations, putting the tablet in the hands of the user and applying learning-by-doing strategies.

3.2. Theme 2. Intuitive user interface may replace extensive instructions

Due to participants lack of vocabulary related to the context they received very little of the planned information about the technology. Despite the minimal spoken and written information, they did not seem to mind, and it did not appear to affect their ability to understand how the technology worked or how they should click and interact with the tablet to achieve what they wanted to do. On the contrary, they felt positive about not having to listen to all the boring information. When presented with the tablet they enthusiastically started exploring. "It doesn't matter that I did not understand what you

tried to say. You know, all those words were not necessary, I have figured it out already, this was very easy".

3.3. Theme 3. Unwillingness to commit to a pre-defined measure of performance

Although participants were positive, some of them eager, to start using social activities online, they did not want to make commitments regarding how much and for how long time they would use the tablet or how often they would engage in activities. If it was mandatory to use a certain frequency or amount, they did not want the service. They argued this by relating to their poor health. "In my condition I can never give promises, because I don't know what I'll be able to do tomorrow". If pressured to perform they felt the tablet would be a burden and a stress factor instead of an asset for pleasure.

4. Discussion

The result from this study contributes with knowledge that there are essential factors to consider when introducing digital social activities to older persons in a home care context. Even though the older persons seem to be literate and digitally experienced it is not safe to assume they understand the context-related words used to label artefacts and functions. In this study lack of vocabulary related to technology meant that part of the information prepared by researchers became useless and therefore excluded. Talking about something that does not make sense was not considered useful.

During the co-creation process preceding this study [11] severe effort was put into creating simple user interface aiming for it to be intuitive. Theme two can be interpreted as a validation that it was a successful approach. Although all ICTs are not appropriate for all individuals it is possible to apply them with consideration to, and strategies designed for, the context and situation for in which it will be applied [10]. Allowing the end user to take control over the activities is a key factor [13].

Theme three revealed that participants were not willing to commit to promise a specified accomplishment. It is important that researchers keep this in mind as quantity of usage is not the same as how useful something is. For older people, having knowledge about possible benefits from using ICT is a factor that gives motivation [14]. The participants in this study all suffered from isolation and feelings of loneliness. Through home care services they were entitled to social activity. Unfortunately in today's society many home care organizations, including those in this study setting, lack resources and face difficulties providing social activities according to clients' needs, thus ICT can provide an alternative [7,9,10].

There are methodological limitations related to the use of fieldnotes as data. One weakness is that they are recorded by an observer and thus subject to (a) memory and (b) possibly the conscious or unconscious bias of the observer [12]. In attempt to minimize this fieldnotes in this project were recorded immediately after leaving the participants.

5. Conclusions

When introducing digital technologies to older persons who are not familiar with ICT it is essential to consider the following factors for a good and sustainable experience for

both the users and the researchers. Be aware, understand, and prepare for, that older persons may not understand common vocabulary related to the context of internet and digital technologies, even though they regularly already use ICT as artefacts and services. Intuitive user interface can compensate for excluded technology-oriented information and inspire older persons to take initiative and explore technology while feeling brave and good about themselves. Do not put pressure on, or make demands, regarding how much older persons with home care need to use ICT when they are introduced to ICT or digital activities. Putting pressure on older persons to perform in a certain way, or to a certain quantity, can cause intended positive benefits to instead be stressful and negative. This may be a challenge for researchers as there is often a demand to measure participants performance in relation to a project or intervention.

References

[1] Donovan NJ, Blazer D. Social isolation and loneliness in older adults: review and commentary of a national academies report. Am J Geriatr Psychiatry. 2020 28(12):1233–1244, doi: 10.1016/j.jagp.2020.08.005

[2] Cacioppo JT, Cacioppo S. Social relationships and health: the toxic effects of perceived social isolation. Soc Personal Psychol Compass. 2014 8(2):58–72, doi: 10.1111/spc3.12087

[3] Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality. Perspect Psychol Sci. 2015 10(2):227–37, doi: 10.1177/1745691614568352

[4] Dickens AP, Richards SH, Greaves CJ, Campell JL. Interventions targeting social isolation in older people: A systematic review. BMC Public Health. 2011 11(1):1-22, doi: 10.1186/1471-2458-11-647

[5] Nilsson I, Häggström Lundevaller E, Fisher AG. The Relationship between Engagement in Leisure Activities and Self-Rated Health in Later Life. Act Adapt Aging. 2017 41(2):175-190, doi: 10.1080/01924788.2017.1306384

[6] Morley JE. Aging in place. J Am Med Dir Assoc. 2012 Jul;13(6):489-92, doi: 10.1016/j.jamda.2012.04.011.

[7] Meagher G, Szebehely M, Mears JE. How institutions matter for job characteristics, quality and experiences: a comparison of home care work for older people in Australia and Sweden. Work Employ. Soc. 2016 30(5):731-749, doi: 10.1177/0950017015625601

[8] Socialtjänstlag. (2001:453) [Social Services Act] Edited by Socialdepartementet. Sweden. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/socialtjanstlag-2001453_sfs-2001-453 Acessed 4 Jan 2023.

[9] Stranz A, Szebehely M. Organizational trends impacting on everyday realities. In Christensen K, Pilling D (Eds.). The Routledge Handbook of Social Care Around the World. London: Routledge. 2018.

[10] Chen YR, Schulz PJ. The Effect of Information Communication Technology Interventions on Reducing Social Isolation in the Elderly: A Systematic Review. J Med Internet Res. 2016 Jan 28;18(1):e18, doi: 10.2196/jmir.4596

[11] Blusi M, Nilsson I, Lindgren H. Older Adults Co-Creating Meaningful Individualized Social Activities Online for Healthy Ageing. Stud Health Technol Inform. 2018 247:775-779. doi:10.3233/978-1-61499-852-5-775775

[12] Braun V, Clarke V. Thematic analysis. In; Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological. American Psychological Association; 2012. p. 57–7, doi:10.1037/13620-004

[13] Lindgren H. Personalisation of Internet-Mediated Activity Support Systems in the Rehabilitation of Older Adults – A Pilot Study, Aime 2009 International Workshop On Personalisation for e-Health. 2009. p. 20-27

[14] Terp R, Kayser L, Lindhardt T. Older Patients' Competence, Preferences, and Attitudes Toward Digital Technology Use: Explorative Study. JMIR Hum Factors. 2021 8(2):e27005, doi:10.2196/27005

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230172

Feasibility of a Virtual Reality App to Promote Pulmonary Rehabilitation

Joseph Finkelstein[a,1], Irena PARVANOVA[a] and Xingye HUO[a]
[a] *Icahn School of Medicine at Mount Sinai, New York, NY, USA*
ORCiD ID: Joseph Finkelstein https://orcid.org/0000-0002-8084-7441

Abstract. One of the major barriers to joining pulmonary rehabilitation (PR) programs is a lack of awareness about its benefits, combined with overall skepticism about regular exercise among COPD patients. Empowering COPD patients with foundational knowledge about PR may potentially facilitate their decision to join a PR program. A virtual reality (VR) app may serve as an engaging and interactive means to deliver PR education; however, the feasibility of this approach in COPD patients is unknown. The goal of this project was to assess the feasibility of VR-based PR education in COPD patients. Using mixed methods design, the feasibility of the VR app was assessed by evaluating its usability, patient acceptance, and its impact on patient knowledge about PR. The results of the usability assessment showed high user acceptance of the VR system and the ability to successfully operate the VR appliances. The use of the VR education app resulted in a statistically significant increase in patient understanding of the main concepts of pulmonary rehabilitation. Further development and evaluation of VR-based systems for patient engagement and empowerment are warranted.

Keywords. Virtual reality, patient empowerment, pulmonary rehabilitation

1. Introduction

Pulmonary rehabilitation (PR) has been shown to significantly improve the quality of life and clinical outcomes in patients with Chronic Obstructive Pulmonary Disease (COPD) [1]. One of the major barriers to joining PR programs is a lack of awareness about its benefits, combined with overall skepticism about regular exercise among COPD patients [2]. Empowering COPD patients with foundational knowledge about PR may potentially facilitate their decision to join a PR program [3]. A recent scoping review concluded that virtual reality (VR) could potentially be instrumental in supporting health education and called for more studies in this field [4]. A VR app may serve as an engaging and interactive means to deliver PR education. However, the feasibility of this approach in COPD patients, who usually are represented by older adults, is unknown. The goal of this project was to assess the feasibility of VR-based PR education in COPD patients.

2. Methods

2.1. System Design

The Virtual Reality app to empower and engage COPD patients in pulmonary rehabilitation (PR) was designed for Oculus Quest 2 system using the Unity

[1] Corresponding Author: Joseph Finkelstein, E-mail: Joseph.Finkelstein@mssm.edu.

Development Platform in our lab. The educational framework was driven by the concepts of adult learning theories previously shown effective for interactive patient education [5]. The system comprised a headset and 2 controllers (Figure 1). To simplify its use, the device was preset to access the PR education app once the headset was mounted. To further simplify the system interface, it was controlled by a single button represented by the trigger of the right or left controller (Figure 2). The educational content was organized into 5 modules: Introduction; Rehabilitation overview; Exercise overview; Rehabilitation benefits; Telerehabilitation overview. Each module starts with a short 2-3-minute instructional video, followed by 4 multiple-choice questions. The user was required to reply to all multiple-choice questions correctly to continue to the next module. If the subject selects a wrong answer, there is an option to go back and select another answer until the questions are answered correctly (Figure 3).

Figure 1. VR Headset.

Figure 2. VR app control.

Figure 3. PR app.

2.2. Study Design

The feasibility of the VR app was assessed by evaluating its usability, patient acceptance, and its impact on patient knowledge about PR. The system testing was carried out in a convenience sample of nine COPD patients. Each participant was instructed to complete a package of pre-test surveys that included socio-demographics, BRIEF Health Literacy Screening Tool [6], and Pre-PR Education Questionnaire [7]. As a next step, the use of the headset and the controllers was explained by a team member. The subject then completed a PR education using the VR app, followed by completing 3 post-task surveys. The post-task surveys contained both quantitative and qualitative data. Post-task questionnaire focused on the completion of 3 main tasks: 1. Start the VR app; 2. Complete the education module; and 3. Answer multiple-choice questions. Next, participants completed a Post-PR Education Questionnaire, an Attitudinal Survey [8], a Heuristic evaluation form [9], and a System Usability Scale (SUS) [10]. A semi-structured qualitative interview was completed at the end of each study visit.

All tasks reflected in the post-task surveys were completed without assistance, however, users could request help anytime. Any requests for assistance were noted. Post-task surveys asked participants to rank each task on a scale of 1 (very difficult) to 5 (very easy), 1 (very unsatisfied) to 5 (very satisfied), 1 (too much time) to 5 (very little time), and 1 (strongly disagree) to 5 (strongly agree). Survey questions for Task 1 included: 1) How difficult or easy was it to start the VR app? 2) How difficult or easy was it to point to the app and start it? 3) How satisfied are you with using this system to complete this task? 4) How would you rate the amount of time it took to complete this task? 5) Is the system visually appealing? 6) Is the system easy to navigate? Survey questions for Task 2 included: 1) How difficult or easy was it to review the content and finish the sections?

2) How difficult or easy was it to advance from one screen to another? 3) How satisfied are you with using this system to complete this task? 4) How would you rate the amount of time it took to complete this task? 5) Is the system visually appealing? 6) Is the system easy to navigate? Lastly, survey questions for Task 3 included: 1) How satisfied are you with using this system to complete this task? 2) How would you rate the amount of time it took to complete this task? 3) Is the system visually appealing? 4) Is the system easy to navigate? These questions were followed by two open-ended questions asking the participant to describe any problems and give additional feedback. The study was approved by the IRB at the Icahn School of Medicine at Mount Sinai.

3. Results

The study sample comprised nine COPD patients, 67% of whom were females, 22% were White, 11% were Asian, and 67% were Black, mean age of 71±7 years old ranging between 59 and 82 years old; none of the study participants used VR in the past. The mean score of the BRIEF Health Literacy Screening Tool was 16.2, indicating a marginal health literacy level that usually requires assistance and is indicative of individuals who struggle with patient education materials.

The mean post-task scores on a scale from 1 to 5 were 4.74, 4.85, and 4.89 for tasks 1 (Start VR app), 2 (Complete the education module), and 3 (Answer multiple-choice questions), respectively. The lowest score for task 1 corresponded to the study participants' feedback, some of whom exhibited difficulties in finding and starting the VR app using VR appliances. Overall, 89% were able to accomplish task 1 without any prompts, and 100% were able to complete tasks 2 and 3 independently. All subjects eventually completed all tasks.

Table 1. Heuristic assessment of the VR app

		N	Minimum	Median	Maximum	Mean
1	The system shows you what's going on and gives you feedback (visibility)	9	4	5	5	**4.89**
2	Language and words make sense (match between system and real world)	9	3	5	5	**4.78**
3	There are clearly marked 'exits', buttons to go back or move forward (control/freedom)	9	4	5	5	**4.89**
4	Words, situations, and actions mean the same thing as elsewhere (consistency)	9	3	5	5	**4.44**
5	There are very few errors, and minimal error-prone conditions (error prevention)	9	3	5	5	**4.78**
6	Instructions are obvious, no need to remember how things work (recognition, not recall)	9	3	5	5	**4.67**
7	The system works for both new and expert users (flexibility, efficiency of use)	9	4	5	5	**4.89**
8	Information is streamlined and relevant (aesthetic / minimalist design)	9	1	5	5	**4.44**
9	Error messages are clear and in plain language (recognize, recover from errors)	9	3	5	5	**4.67**
10	Help is available, searchable, and relevant (help and documentation)	9	4	5	5	**4.89**

The results of the heuristic evaluation of the user interface are presented in Table 1. The mean score was 4.7 (on a scale between 1 and 5), indicating sufficient usability of the VR app, with the consistency dimension receiving the lowest score and visibility and flexibility – the highest. The mean System Usability Score (SUS) was 95.8. With the maximum SUS score of 100, a score of 95.8 indicates a high usability assessment that is typically indicative of people who love an app or site and will recommend it to their friends. The pre-post comparison of PR knowledge demonstrated a statistically significant increase in PR knowledge after using the VR app based on the paired t-test (an increase of the mean score from 7.2 to 7.9; $p<0.04$).

Table 2. Concept map

CONTENT		INTERFACE		PROCESS	
Facilitators	**Barriers**	**Facilitators**	**Barriers**	**Facilitators**	**Barriers**
Clear	Low volume during the last video	Amazing	Additional information needed	Accessible	Lack of additional COPD information
Different		Concise	Clearer directions needed	Assists with focus	Lack of exercises
Easy to see		Convenient	Heavy headset	Beneficial	Lack of music
Enjoyable		Easy to operate	Lack knowledge about the location of the volume button	Good idea	Lack of research information
Flawless	None	Engaging	Lack of manual	Interesting	None
Interactive		Excellent	None	Short	
Relevant information		Fun		Useful	
Self-explanatory		Good experience			
Three-dimensional		Good learning tool	Tight headset	Understandable	
		Intuative			

The semi-structured qualitative interviews were analyzed using thematic analysis [11] which showed high acceptance of the VR system content, interface, and process. A number of valuable suggestions on how to improve the system were provided by the patients. Based on the qualitative analysis results, a concept map [12] was built to represent facilitators and barriers across three major themes: app content, app interface, and app process (Table 2).

4. Discussion

An interactive virtual reality app for COPD patient education and engagement in pulmonary rehabilitation was tested in a convenience sample of VR-naïve COPD patients. This is the first study testing a VR app for PR education. The patients demonstrated a high interest in using the interactive VR format for health education and personal empowerment. The results of the usability assessment based on the cognitive walkthrough and heuristic assessment of the user interface showed high user acceptance of the VR system and the ability to successfully operate the VR appliances. The use of the VR education app resulted in a statistically significant increase in patient understanding of the main concepts of pulmonary rehabilitation. Semi-structured qualitative interviews uncovered barriers and facilitators of VR use for patient engagement and empowerment. Our results concur with the previous studies that

demonstrated that avatar-based or mobile interactive platforms using concepts from adult learning theories could significantly increase disease-specific knowledge in patients with low health literacy and limited computer skills [5,13].

5. Conclusion

The interactive VR app for patient education and empowerment was well accepted by COPD patients. The use of the VR app resulted in a statistically significant increase in patient knowledge about pulmonary rehabilitation. Further development and evaluation of VR-based systems for patient engagement and empowerment are warranted.

6. Acknowledgements

The project was funded by NIH grant R33HL143317.

References

[1] Fleg JL, Keteyian SJ, Peterson PN, et al. Increasing Use of Cardiac and Pulmonary Rehabilitation in Traditional and Community Settings: Opportunities to Reduce Health Care Disparities. J Cardiopulm Rehabil Prev. 2020 Nov;40(6):350-355.
[2] Shero ST, Benzo R, Cooper LS, et al. Update on RFA Increasing Use of Cardiac and Pulmonary Rehabilitation in Traditional and Community Settings NIH-Funded Trials: Addressing Clinical Trial Challenges Presented by The Covid-19 Pandemic. J Cardiopulm Rehabil Prev. 2022 Jan 1;42(1):10-14.
[3] McNamara RJ, Dale M, McKeough ZJ. Innovative strategies to improve the reach and engagement in pulmonary rehabilitation. J Thorac Dis. 2019 Oct;11(Suppl 17):S2192-S2199.
[4] van der Kruk SR, Zielinski R, MacDougall H, et al. Virtual reality as a patient education tool in healthcare: A scoping review. Patient Educ Couns. 2022 Jul;105(7):1928-1942.
[5] Finkelstein J, Lapshin O, Cha E. Feasibility of promoting smoking cessation among methadone users using multimedia computer-assisted education. J Med Internet Res. 2008 Nov 3;10(5):e33.
[6] Haun, J., Luther, S., Dodd, V., & Donaldson, P. (2012). Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun. 2012;17 Suppl 3:141-59.
[7] Castro H, Hise M, Finkelstein J. A comparison of two models of web-based education in older adults. AMIA Annu Symp Proc. 2005;2005:914.
[8] Barron J, Bedra M, Wood J, et al. Exploring three perspectives on feasibility of a patient portal for older adults. Stud Health Technol Inform. 2014;202:181-4.
[9] Robins D, Brody R, Parvanova I, et al. Cognitive Testing of an Electronic Consent Platform: Researcher Perspectives. Stud Health Technol Inform. 2021 Dec 15;284:457-462.
[10] Brooke, J. SUS: a “quick and dirty” usability scale. Digital Equipment Corporation, 1986.
[11] Ritchie J. and Spencer L. Qualitative data analysis for applied policy research. Analyzing Qualitative Data. 1994; 9:1-22.
[12] Johnson BD, Dunlap E, Benoit E. Organizing “mountains of words” for data analysis, both qualitative and quantitative. Subst Use Misuse. 2010;45(5):648-70.
[13] Bedra M, Wick E, Brotman D, et al. Avatar-based interactive ileostomy education in hospitalized patients. Stud Health Technol Inform. 2013;190:83-5.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230173

A Web-Based Public Health Intervention for Addressing Vaccine Misinformation: Analysis of Learner Engagement and Shift in Hesitancy to Vaccinate

Leigh POWELL[a,1], Radwa NOUR[a] , Hanan AL SUWAIDI[b] and Nabil ZARY[a]

[a] *Institute for Excellence in Health Professions Education, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE*
[b] *College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE*
ORCiD ID: Leigh Powell https://orcid.org/0000-0001-6714-7371, Radwa Nour https://orcid.org/0000-0001-8133-1562, Hanan Al Suwaidi https://orcid.org/0000-0002-1261-4224, Nabil Zary https://orcid.org/0000-0001-8999-6999

Abstract. Web-based public health interventions can be a useful tool for disseminating evidence-based information to the public. However, completion rates are traditionally low, and misinformation often travels at a faster pace than evidence-based sources. This study describes the design of a web-based public health intervention to address COVID-19 vaccine hesitancy. A quasi-experimental approach was used in which a validated instrument, the Adult Vaccine Hesitancy Survey, was given to learners both pre and post intervention to observe any change in attitude towards vaccination. Our pilot observed a small positive shift in vaccine hesitancy and experienced higher than average completion rates. By integrating motivational learning design into public health interventions we increase the likelihood that learners finish the entire intervention, creating greater chance for positive behavior change.

Keywords. Health communication, health intervention, learning analytics, vaccine hesitancy, public health, COVID-19

1. Introduction

Public health education interventions (HEIs) are those which aim to improve both access and delivery of information to address social determinants of health by empowering behavior change [1]. HEIs are a means of promoting health communication, raising awareness about public health concerns, correcting misinformation, and promoting behavior change [2]. Previous studies have shown that HEIs have aided in the prevention and control of communicable diseases, such as SARS (severe acute respiratory syndrome) [3] and MERS (Middle East respiratory syndrome) [4]. Unsurprisingly, web-

[1] Corresponding Author: Leigh Powell, E-mail: leigh.powell@mbru.ac.ae.

based HEIs have become more prevalent, owing to their ability to overcome physical barriers and reach broader populations online.

Misinformation about the safety of COVID-19 vaccines spread virally online [5]. Health misinformation has been defined as "a health-related claim that is based on anecdotal evidence, false, or misleading owing to the lack of existing scientific knowledge" [6,7]. Vaccine misinformation led to a rise in vaccine hesitancy, "a delay in acceptance or refusal of vaccination despite the availability of vaccination services" [8]. Web-based HEIs may have a positive impact on vaccine hesitancy, reaching the public online where the misinformation is being spread.

Web-based HEIs, such as massive open online courses (MOOCs), come with many challenges, including a wide variability in results, effectiveness, completion rates, and attrition rates. Reasons for this variability may include a lack of attention in design for learner motivation and interactivity [9]. Designing HEIs with motivational learning design strategies is a way of engaging learners to achieve learning objectives [10].

2. Objectives

This study explores the development and engagement in a web-based, HEI designed to influence attitudinal change towards COVID-19 vaccine hesitancy. This study explored two research questions: 1) What kinds of learner engagement did we observe in the intervention? 2) What changes are observed in a participant's vaccine hesitancy status as a result of the intervention, as determined by a pre-post validated survey tool?

3. Methods

The released version of our HEI is titled, Level Up! Fight the COVID-19 Misinformation Pandemic. Level Up (LU) is a self-paced HEI distributed for free on our platform MBRU Learn, an online rapid course development platform. LU consists of three stages of learning with associated learning objectives (Table 1) and takes no longer than 30 minutes to complete.

Table 1. Topics and Learning Objectives for the pilot release of Level Up

Stage	Topic	Learning Objectives
1	Understanding and counteracting misinformation	Describe the negative impact that misinformation has on the pandemic. Evaluate information to avoid misinformation traps
2	COVID-19 transmission and prevention	Apply strategies to reduce the risk of contracting and spreading COVID-19
3	The science of vaccinations	Recognize the science behind vaccinations and how they work to protect public health.

LU uses the design an 8-bit, video-game like experience in which the learner follows an avatar as he grows in strength and abilities in each stage. Gamification principles were integrated into the design in which learners' complete knowledge checks to collect objects (Figure 1), grow in skills and rank, and work towards mastery to 'Beat the Boss'

(Figure 2) and achieve the final rank of Legend. Learners have unlimited attempts to pass knowledge checks and must achieve 100% to progress. Throughout each stage a variety of media and interactive content is used to engage the learners, including animated videos, motion graphics, HTML5 games, audio and text.

Figure 1. Screenshot from Level Up—collection of objects.

Figure 2. Screenshot from Level Up— Beat the Boss challenge.

The aim of LU is to address misinformation related to COVID-19 vaccines and measure whether a change in vaccine hesitancy occurred. This study used a quasi-experimental design in which the Adult Vaccine Hesitancy Scale (aVHS), a validated instrument, was given to learners both pre and post intervention. The aVHS contains 10-items scored on a Likert scale. All scores and cutoffs were adopted by following the methodology of the research team developing the scale [11]. Statistical analysis software was used for analysis . Analytics data was exported from MBRU Learn and cleaned to enable an exploratory analysis of learner engagement and retention, two of the most common measurements used to understand and improve web-based HEIs [12].

4. Results

A total of 641 learners enrolled in Level Up, with a total of 408 learners completing the entire intervention. 88 learners dropped off immediately after enrollment and did not attempt any content. A detailed breakdown of learners at each stage of Level Up is provided in Table 2. The biggest loss of learners occurred at Stage 1, with 68 learners attempting the stage but not finishing. Analysis revealed that the quiz in Stage 1 was attempted an average of 5.5 times per learner, with question 2 being the question that was most often missed, with 2002 incorrect attempts.

Table 2. Breakdown of learners at each stage and section of Level Up (N=641) *cumulative by stage

	Completed zero content *	Attempted content	Total non-completers *	Exited between stages	Completed whole stage	% completed
Pre survey	88	38	126	NA	515	80.34%
Stage 1	135	68	203	9	438	68.33%
Stage 2	204	8	212	1	429	66.93%
Stage 3	212	9	221	0	420	65.52%
Beat the Boss	221	11	232	0	409	63.81%
Post survey	232	1	233	0	408	63.65%

A total of 300 learners completed both the pre and post aVHS. A total score of the aVHS was calculated for each pre and posttest. Learner shift from vaccine hesitancy pre-post is described in Table 3, with 35 hesitant learners and 265 non-hesitant learners after pre-test. We observed a 5.4% decrease in hesitancy after the post-test, with the proportion of vaccine hesitant learners dropping from 11.7% (n=35) at pre-test to 6.3% (n=19) after the post-test. Only 3 learners, or 1%, of our 265 non-hesitant learners at pre-test shifted from non-hesitant to hesitant. Chi squared (χ^2) showed significant results p<.001. The mean score for the post-test was lower than the pre-test (mean pre-test 18.6, SD=5.21, mean post-test 16.62, SD= 5.21) indicating a reduction in hesitancy. The difference in means was 1.98, 95% CI (1.48-2.48), t=7.8273, p< .0001.

Table 3. Hesitancy changes from pre to post test (N=300)

			Hesitant (post)	Not Hesitant (post)
Hesitant (pre)	35			
		After post	16	19
Not Hesitant (pre)	265			
		After post	3	263

5. Discussion

Level Up was developed to understand if a web-based HEI, developed using motivational learning design, could make an impact on leaners hesitancy to vaccinate. Our pre-post test results showed some impact on vaccine hesitancy with 5.34% of previously hesitant learners shifting into not hesitant status after completing the intervention and post-test. A backward shift was also observed in 3 participants (1%) who went from not hesitant to hesitant. A larger population size is necessary to draw further conclusions. Completion rates for LU (64%) were higher as compared to online web-based interventions, such as MOOCS, which typically experience completion rates from .7% to 52% [13,14], an indicator that our motivational learning design strategies were effective. In their systematic review of papers using gamification for MOOCs, de Freitas and da Silva [12] found greater participation rates from learners when gamification principles were used as reflected by the time spent on MOOC platforms, the number of learners completing end-of-course evaluations, and the number of tasks and lessons completed [12]. Our loss of 68 learners in Stage 1 appears to be a result of a difficult choose-all-that-apply (CATA) question on the Stage 1 quiz. Visual placement of instructions is quite important for CATA questions [15] which is not changeable within MBRU Learn, so future work will see this assessment reworked.

6. Limitations

We are limited by our pilot population size and more participants are needed to draw further conclusions the impact on vaccine hesitancy. The majority of our population were also fully vaccinated so we would benefit from a greater diversity of learners. We did not publicly advertise for enrollments so learners in this are those who already follow MBRU Learn and therefore might be more motivated to complete our interventions. We

are also limited by our platform from seeing more granularity when it comes learner activity. The intervention is only available in English and to those who have an Internet connection.

7. Acknowledgments

The study was approved by the MBRU institutional review board in October 2021 (submission number: MBRU IRB-2021-68). This work was supported in part by the Al Jalila Foundation and funded by a Pfizer Independent Medical Education Grant (agreement number: 67504787).

References

[1] Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promotion International. 2000 Sep 1;15(3):259–67.
[2] Abiodun OA, Olu-Abiodun OO, Sotunsa JO, Oluwole FA. Impact of health education intervention on knowledge and perception of cervical cancer and cervical screening uptake among adult women in rural communities in Nigeria. BMC Public Health. 2014 Aug 7;14(1):814.
[3] Li W, Liao J, Li Q, Baskota M, Wang X, Tang Y, et al. Public health education for parents during the outbreak of COVID-19: a rapid review. Ann Transl Med. 2020 May;8(10):628.
[4] Alqahtani AS, Wiley KE, Mushta SM, Yamazaki K, BinDhim NF, Heywood AE, et al. Association between Australian Hajj Pilgrims' awareness of MERS-CoV, and their compliance with preventive measures and exposure to camels. J Travel Med. 2016 May;23(5):taw046.
[5] Lawhorne L, Swedlund S. The Infodemic within the Pandemic. J Am Med Dir Assoc. 2021;22(3):517–8.
[6] Chou WYS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA. 2018 Dec 18;320(23):2417–8.
[7] Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. Journal of Medical Internet Research. 2021 Jan 20;23(1):e17187.
[8] MacDonald NE. Vaccine hesitancy: Definition, scope and determinants. Vaccine. 2015 Aug 14;33(34):4161–4.
[9] Khalil H, Ebner M. MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review. In Association for the Advancement of Computing in Education (AACE); 2014 [cited 2023 Feb 27]. p. 1305–13. Available from: https://www.learntechlib.org/primary/p/147656/
[10] Pittenger A, Doering A. Influence of motivational design on completion rates in online self-study pharmacy-content courses. Distance Education. 2010 Nov;31(3):275–93.
[11] Akel KB, Masters NB, Shih SF, Lu Y, Wagner AL. Modification of a vaccine hesitancy scale for use in adult vaccinations in the United States and China. Human Vaccines & Immunotherapeutics. 2021 Mar 26;1–8.
[12] Jarnac de Freitas M, Mira da Silva M. Systematic literature review about gamification in MOOCs. Open Learning: The Journal of Open, Distance and e-Learning. 2020 Aug 11;1–23.
[13] Antonaci A, Klemke R, Lataster J, Kreijns K, Specht M. Gamification of MOOCs Adopting Social Presence and Sense of Community to Increase User's Engagement: An Experimental Study. In: Scheffel M, Broisin J, Pammer-Schindler V, Ioannou A, Schneider J, editors. Transforming Learning with Meaningful Technologies. Cham: Springer International Publishing; 2019. p. 172–86. (Lecture Notes in Computer Science).
[14] Jordan K. Massive Open Online Course Completion Rates Revisited: Assessment, Length and Attrition. irrodl. 2015;16(3):341–58.
[15] Ares G, Etchemendy E, Antúnez L, Vidal L, Giménez A, Jaeger SR. Visual attention by consumers to check-all-that-apply questions: Insights to support methodological development. Food Quality and Preference. 2014 Mar 1;32:210–20.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230174

Danish Citizens' Expectations for mHealth Prescription

Victor Vadmand JENSEN[a,1] Tom Høg SØRENSEN[b] Alice Ørsted KRISTENSEN[b] and Jeppe ERIKSEN[c]
[a] *Department of Computer Science, Aalborg University, Denmark*
[b] *MedCom, Odense, Denmark*
[c] *Department of Planning, Aalborg University, Denmark*
ORCiD ID: Victor Vadmand Jensen https://orcid.org/0000-0003-4270-7891, Jeppe Eriksen https://orcid.org/0000-0003-0751-4437

Abstract. Use of mHealth in Denmark is growing, and prescription structures for mHealth apps are a political goal. In this pilot survey study, respondents generally perceive their mHealth use as beneficial, which correlates with their frequency of use. Willingness to substitute traditional treatments for prescribed mHealth varies based on type of substituted treatment.

Keywords. mHealth, smartphone apps, mHealth prescription, citizen expectation

1. Introduction

The Danish government's strategy for life science focuses on development of prescription structures for quality-assured mHealth [1], a development supported by increased use of mobile apps among Danish citizens [2]. However, many smartphone apps lack content appropriate for prescribing them to specific patient groups [3]. This short communication examines Danish citizens' expectations for mHealth prescription to develop considerations for such prescription.

2. Methods

A questionnaire was developed covering: (i) demographic data, (ii) use of mHealth, and (iii) expectations for mHealth prescription from physicians. Likert-scale questions were used for measuring (ii) using perceived usefulness (PU) [4], while (iii) was measured with performance expectancy (PE) [4] and substitutive use (SUB) [5]. Cronbach's alpha was calculated for PU ($\alpha = 0.86$), PE ($\alpha = 0.82$) and SUB ($\alpha = 0.73$). The opt-in Google Forms questionnaire was distributed via QR codes at the event The People's Meeting, LinkedIn, and Facebook. No personally identifiable information was collected.

[1] Corresponding Author: Victor Vadmand JENSEN; E-mail: vjense17@student.aau.dk.

3. Results

67 respondents participated in the survey, 20 men (29.9%) and 47 women (70.1%), with 53 respondents (79.1%) aged < 60 years. 56 respondents (83.6%) had a higher education ≥ 3 years in length. Regarding questions on PU of mHealth, mean PU for respondents using mHealth: once a month or less was 3.49 (n = 17); 2-3 times a month was 3.83 (n = 10); 1-6 times a week was 4.27 (n = 25); once a day or more was 4.33 (n = 15).

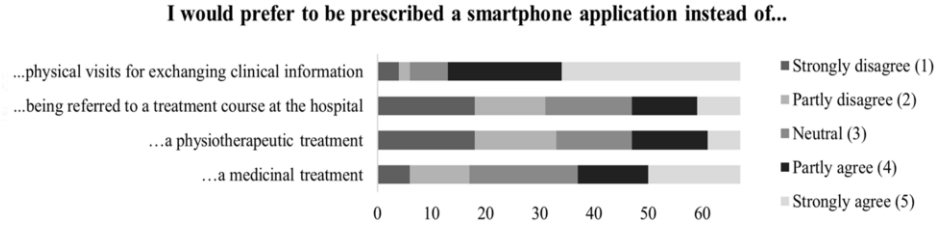


Figure 1. Respondents' willingness to substitute traditional treatments for prescribed mHealth (SUB).

Figure 1 show that 54/67 and 30/67 of respondents prefer substitution of physical information exchange and medicinal treatment, respectively, with mHealth.

4. Discussion

Results raise questions about use cases of mHealth prescription which point to substitution of pain-relief medicine and clinical information exchange, the latter possibly via telemedicine. Further, possible connection between frequency of use and PU suggests a possibility of prescribing mHealth often used by citizens. The sample is biased toward younger respondents with longer educations, requiring further investigation for disadvantaged citizens to benefit from mHealth prescription.

Future research should develop indicators for assessment of prescribed mHealth apps already used among citizens. Researchers should also scrutinize the appropriate use cases for mHealth prescription as a pragmatic counter to mHealth prescription hype.

References

[1] Ministry of Industry, Business and Financial Affairs. Agreement on a strategy for life science [Internet]. Place of publication: Ministry of Industry, Business and Financial Affairs; 2021 May [cited 2023 February 22]. Available from: https://em.dk/media/14236/agreement-on-a-strategy-for-life-science.pdf.

[2] Eriksen J, Tornbjerg Eriksen K, Bertelsen PS. Citizens' use of Health Information Technology between 2013-2021 in Denmark: A longitudinal study. In The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Linköping Electronic Conference Proceedings 187. 2022. p. 128-135 doi: 10.3384/ecp187F

[3] vPayo RM, Harris J, Armes J. Prescribing fitness apps for people with cancer: a preliminary assessment of content and quality of commercially available apps. J Cancer Surviv. 2019 Jun;13(3):397-405. doi: 10.1007/s11764-019-00760-2.

[4] Alam M, Hu W, Barua Z. Using the UTAUT Model to Determine Factors Affecting Acceptance and Use of Mobile Health (mHealth) Services in Bangladesh. Journal of Studies in Social Science. 2018;17(2): 137-172.

[5] Rai A, Chen L, Pye J, Baird A. Understanding Determinants of Consumer Mobile Health Usage Intentions, Assimilation, and Channel Preferences. J Med Internet Res. 2013 Aug; 15(8): e149. doi: 10.2196/jmir.2635

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230175

First Contact, First Learnings - Nursing Staff Approaching Robotics in Health Care

Jörn KRÜCKEBERG[a,1], Daniel Peter BEUME[b], Ronny KLAWUNN[c], Jens MÜLLER[a], Leif GRIGA[d], Lena WINKELHAKE[d], Lukas EICH[d], Nina KEMPE[d] and Esma ALBAYRAK[d]
[a] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig - Institute of Technology and Hannover Medical School, Germany*
[b] *Hochschule Hannover - University of Applied Sciences and Arts, Germany*
[c] *Institute for Epidemiology, Social Medicine and Health Systems Research, Hannover Medical School, Germany*
[d] *Robokind - Robotics for Mankind Foundation, Hannover, Germany*

Abstract. Systems for service and assistance robotics become relevant in nursing care. Workshops with target user groups can support the reflection and identification of scenarios for the use of robotic systems.

Keywords. Nursing care, Nursing Robotic, Participation

1. Introduction

The current discussion about technical support, especially robotic systems in nursing, requires a comprehensive exchange between all parties involved in introducing technical products. This exchange is needed to integrate these products in reasonable fields of activities in nursing [1].

The research project Centre of Implementing Nursing Care Innovations (PPZ Hannover) aims to integrate innovative technologies such as robotic systems to support nursing staff and to improve patient care. Emotional robotics - with systems such as the Paro seal, JustoCat or a UV disinfection robot - has already been tested or discussed as part of the project´s participatory implementation concept [2]. Now service and assistance functions for support in everyday care are also coming into focus with robot systems such as "Lio" [3, 4]. This contribution presents initial results of a workshop for nurses on robotics for care delivery.

2. Methods

Together with the Robokind Foundation a workshop concept was developed that starts with a one-week e-learning phase before the workshop date. The workshop itself

[1] Corresponding Author: Jörn Krückeberg, Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany; Email: Joern.Krueckeberg@plri.de.

integrates a survey on participants´ occupational backgrounds and experiences and in reflexive group phases:

- discusses questions about the participants´ attitudes and perceptions towards robotic systems in care,
- introduces the current state of technical development of robots with different example systems and
- defines tasks for nursing robots and demands on the technical design on the basis of identified fields of application in everyday nursing care.

3. Results

Three workshop dates at the end of 2022 were filled to capacity with 12 participants each. The initial results of the discussion indicate that the participants predominantly recommend coexistent and cooperative fields of activity in which robots can provide support, for example, in service tasks such as serving meals. Less activities that are closer to the patient or the body are seen in robotic assistance with measurements of vital data.

Furthermore, the results show an ambivalent attitude of the participants with regard to the question whether robotic systems with reduced specific tasks and functions are more helpful than multifunctional, more complex systems whose development and operation is more complex.

4. Conclusion

The results of the workshop underline the suitability of participatory processes that involve relevant target groups in nursing in the reflection and design of scenarios for the use of robotic systems in nursing practice. The rather complex workshop concept is contrasted by results that can contribute to the identification of relief potentials and the technical development of practical robotic systems in nursing by means of constructive-critical expertise.

One requirement for achieving such results is an equal level of knowledge among all participants with regard to technical possibilities, actual needs, data protection regulations and ethical implications of the use of robotics. Therefore, further educational offers on the topic of robotics in nursing care are essential.

References

[1] Paluch R, Müller C, Garthaus M, Hülsken-Giesler M, Stachura E (2022) Meeting report: „Robotische Systeme für die Pflege“. Symposium, 2022 (online). TATuP [Internet]. 2022 Jul. 18 [cited 2022 Dec. 13];31(2):78-9. Available from: https://www.tatup.de/index.php/tatup/article/view/6976

[2] Klawunn R, Dierks M D, Krückeberg J, & Hechtel N (2021) Partizipation von Pflegefachpersonen in Technikauswahl und -einführung. Public Health Forum 29(3): 260–264. https://doi.org/10.1515/pubhef-2021-0065

[3] Nieto Agraz C, Pfingsthorn M, Gliesche P, Eichelberg M and Hein A (2022) A Survey of Robotic Systems for Nursing Care. Front. Robot. AI 9:832248. doi: 10.3389/frobt.2022.832248

[4] Lio - F&P Robotics [Internet] Glattbrugg, Switzerland: F&P Robotics AG; [accessed 2023 March 03]. Available from: https://www.fp-robotics.com/en/lio/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230176

Access to Development Opportunities in Biomedical and Health Informatics

Sidsel VILLUMSEN[a,1], Rikke Amalie Agergaard JENSEN[b,c], Klaus NIELSEN[a], Ouafa RIAN[d] and Charlotte JONASSON[a]
[a] *Dept. of Psychology, Aarhus University, Denmark*
[b] *Dept. of Regional Health Services Research, University of Southern Denmark, Denmark*
[c] *Mental Health Services Region of Southern Denmark, Denmark*
[d] *Koncern HR, Udvikling, Central Region Denmark, Aarhus, Denmark*

Abstract. In between users and trained informaticians, we find a group of people carrying out important work in implementing and further developing health information technology, without access to formal biomedical and health informatics (BMHI) training. Study findings show what is required of novices in BMHI to gain access to communities of practice through which expertise can be developed.

Keywords. competencies, healthcare workforce, communities of practice

1. Introduction

With the increase of healthcare and information technologies (HIT) in public healthcare follows a widely recognized need for correspondingly training the healthcare workforce to be capable of adopting and adapting new technologies [1]. As they are not, nor should be, informaticians, they do not have access to formal BMHI training, leaving them to seek out the necessary skills and expertise in their professional network. We suggest a theoretical framework of communities of practice [2, 3], where novices develop skills through participating in actual 'situated' work practices with experts. This study explores how novices in BMHI gain access to such communities of practice.

2. Methods

The study was conducted in the secondary public healthcare sector in a Scandinavian country, focusing on employees and managers' participation in department-local HIT projects. Data originates from participant observations (throughout 2020-2022) of project meetings, collaborative meetings, and everyday work as well as semi-structured

[1] Corresponding Author: Sidsel Villumsen, Department of Psychology and Behavioural Sciences, Aarhus University, Bartholins alle 11, 8000 Aarhus C, Denmark. Email: sidvil@psy.au.dk, phone: +45 93522753.

The study is part of the LEARNGOV project funded by Independent Research Fund Denmark, case number: 1024-00018B. Grant holder associate professor Charlotte Jonasson.

interviews (n=29) with project participants (healthcare workers), managers, and informaticians. Data from participant observations and interviews was coded according to thematic analysis [4].

3. Results

Findings highlight three aspects of gaining access in the planning and implementing of HIT: 1) knowing when expert help was needed, 2) knowing which kind of expertise was needed, and 3) insisting on continued expert relations. It was complicated for the novices to plan and implement HIT. They would often encounter various obstacles spanning from technical problems to finding time and resources. The novices had to negotiate whether they could overcome obstacles themselves or if they had to reach out to experts. They had to show great patience when figuring out how to gain access to experts, who were able and willing to help them. Although this took time, they gained important understanding of BMHI as well as of implementing HIT in their own, complex organization. Even more experienced novices, who knew when and whose expertise was needed, had to insist on continued relations to shifting experts in complex, ever-changing communities of practice. These three aspects of gaining access were required to implement HIT. At the same time, through this process of gaining access, the novices widened their knowledge of and situated training in BMHI.

4. Discussion and Conclusion

It is no secret that planning and implementing HIT requires access to the right experts. However, our findings show that this is not as easy as it may sound, and furthermore, it is highly context sensitive. Knowing *when* to ask for help, *where* to find it, and *keeping* this access to expertise available is central to becoming part of a community of practice through which expertise in HIT and BMHI is developed.

Because the needs are diverse, depending on the specific project, and the HIT landscape is everchanging, approaching competency development in the wider healthcare workforce through life-long, situated learning could be advantageous [1]. We call for a strengthened focus on gaining legitimate access to participate in projects with relevant experts [3] as it is through the everyday work with digitalization projects and experts in various areas that the healthcare workforce can develop expertise in BMHI.

References

[1] Knudsen C, Villumsen S, Krejberg L, and Nøhr C. The Last Mile Problem and Beyond in HIT – The Role of Context-Sensitive Digital Integrators, in: *Studies in health technology and informatics*, IOS Press, 2022, pp. 97-101.

[2] Wenger E. Communities of practice: Learning, meaning, and identity. Cambridge university press; 1999 Sep 28..

[3] Lave J, Wenger E. Situated learning: Legitimate peripheral participation. Cambridge university press; 1991 Sep 27.

[4] Miles MB, Huberman AM, Saldana J. Qualitative data analysis 3rd Edition: Source book of Bew Methods. Baverly Hills: SAGE Publications Inc. 2014.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230177

Automated Monitoring Reports of the Activity of the French National Professional Suicide Prevention Helpline

Antoine LAMER[a,b,c,1], Chloé SAINT-DIZIER[a,b], Emile FARES[a], Christophe DEBIEN[a], Elise CLEVA[a], Marielle WHATELET[a] and Charles-Edouard NOTREDAME[a]
[a]*F2RSM Psy - Fédération régionale de recherche en psychiatrie et santé mentale Hauts-de-France, F-59350, Saint-André-Lez-Lille, France*
[b]*Univ. Lille, Faculté Ingénierie et Management de la Santé, F-59000, Lille, France*
[c]*Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de santé et des Pratiques médicales, F-59000 Lille, France*

Abstract. The French Professional Suicidal Helpline 3114 was launched on October 1st, 2021. The objective of this study was to implement automated reports of the activity of the suicidal helpline. We developed automated reports and presentations with Rmarkdown. Two formats were developed, national reports to present for a funding agency and regional reports for each calling center. These reports fulfill a critical need to adjust call distribution patterns, identify problems, adjust communication across the territory and ensure that 3114 is delivering the service it is supposed to provide.

Keywords. Data visualization, visual analytics, suicide, psychiatry, helpline

1. Introduction

The 3114, a French National Suicide Prevention Helpline, has been launched on the 1st October, 2021. The 3114 is a professional, free and confidential hotline [1]. Health professionals in fifteen different regional centers handle the daily calls. Indicators about the number of calls received, the rates of answered calls, their durations and their distributions by center were needed to assess the activity of the helpline. In healthcare, dashboards and automated reports are implemented to improve the decision-making process [2]. The objective of this study was to generate automated reports of the activity of the suicidal helpline.

2. Methods

We developed automated reports and presentations with Rmarkdown (R packages knitr, version 1.39 and rmarkdown, version 2.14), a technology implemented with R

[1] Corresponding Author, Antoine Lamer, F2RSM Psy, Lille University, 2 place de Verdun, F-59000, Lille, France; E-mail: antoine.lamer@f2rsmpsy.fr.

(version 4.1.3). Two reports formats were developed, a national report for presentation to the French Ministry of Health and Prevention and a regional for each calling center. They contain common indicators, such as the number of received calls, answered call and the response rates. The projection of expected activity for the next two months is performed with an ARIMA model (R package forecast, version 8.20). Reports could also contain specific studies on the impact of an event on the activity of the helpline.

3. Results

The hotline received 17,493 (15,436;20,659) calls by month between its launch and November 31st, 2022. 3114's activity is steadily increasing with approximately 871.6 additional calls per month. Most calls are made between 3:00 pm and 10:00 pm (40.0%). For answered calls, the median (Q1;Q3) conversation time was 9.0 (2.1;20.0) minutes. Three centers handle most of calls (39.0%) and 29.1% of calls were received during nights or weekends. The figure 1 displays the daily evolution of incoming calls.

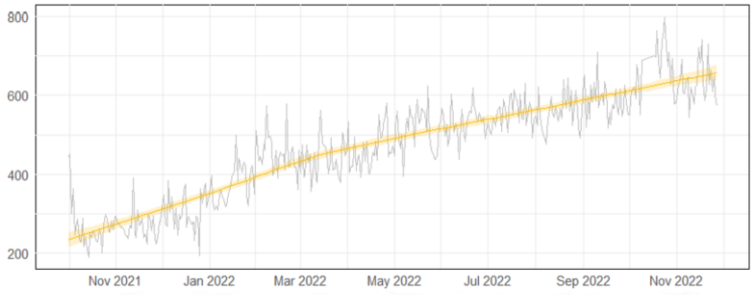


Figure 1. Daily evolution of incoming calls of the French Professional Suicidal Helpline 3114 since its launching on the October 1st, 2021.

4. Discussion and Conclusion

We implemented automated reports to support the monitoring of the activity of the French suicidal helpline, 3114. These reports fulfill a critical need to adjust call distribution patterns, communication across the territory and ensure that 3114 is delivering the service it is supposed to provide. They help improving the response rate by spreading more parsimoniously incoming calls in time and space and by more rationally allocating human resources.

References

[1] Notredame CE, Wathelet M, Morgiève M, Grandgenèvre P, Debien C, Mannoni C, Pauwels N, Ducrocq F, Leaune E, Binder P, Berrouiguet S. The 3114: A new professional helpline to swing the French suicide prevention in a new paradigm. European psychiatry. 2022 Oct 7:1-1. doi: 10.1192/j.eurpsy.2022.2318.

[2] Caban JJ, Gotz D. Visual analytics in healthcare–opportunities and research challenges. Journal of the American Medical Informatics Association. 2015 Mar 1;22(2):260-2. doi: 10.1093/jamia/ocv006.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230178

Health Informatics Training Program in Low and Middle Income Countries

Armine LULEJIAN,[a,1] Lynn KYSH[b], Tamara GALOYAN[c] and James DICKHONER[b]
[a] *Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*
[b] *Innovation Studio, Children's Hospital Los Angeles, Los Angeles, CA, USA*
[c] *School of Education, Drexel University, Philadelphia, Pennsylvania, USA*
ORCiD ID: Lynn Kysh https://orcid.org/0000-0003-1886-0546
ORCiD ID: James Dickhoner https://orcid.org/0000-0002-5727-9300

Abstract. We developed the first health informatics training program in Armenia and in the Caucasus region. The training program consists of four educational pillars, including a bootcamp, an individualized training program, a capstone, and a scholarly project. We conducted surveys and qualitative interviews to evaluate the training program. With trending positive results we acknowledge that it is important to understand the landscape of health informatics and conduct needs assessment prior to establishing such a training program in an LMIC.

Keywords. LMICs, clinical informatics, training, Armenia

1. Introduction

Health information technologies including Electronic Health Records and telehealth are being widely implemented in Low- and Middle-Income Countries (LMIC), especially since the COVID-19 pandemic [1,2]. Many LMICs, including the Republic of Armenia, are in urgent need for training of health informaticians to support best practices and innovation in the domains of clinical care, public health policy, translational research and public health informatics [3]. Armenia possesses the building blocks for a national electronic health record system that can leapfrog the systems in the US and Europe if there is a workforce to build the necessary infrastructure.

2. Methods

Our inaugural training program included four fellows: two physicians and two computer engineers. Over three months, fellows participated in a twenty module bootcamp modeled after AMIA's 10x10 modules, which introduced the field of clinical informatics [4]. Our fellows were placed in the Ministry of Health of the Republic of Armenia, ARMED (Armenia's national EHR vendor), National Institute of Health of the Republic of Armenia, and the National Pulmonary Center. Each fellow will then complete two projects over the next twelve months that are relevant to their current roles and advance

[1] Corresponding Author: Armine Lulejian, E-mail: alulejia@usc.edu

health informatics in Armenia. Each project has two components, (1) a practical implementation, and (2) a scholarly evaluation or research project. To enable each fellow to complete their projects, we are providing supplementary coursework and mentors with relevant expertise.

We conducted four semi-structured qualitative interviews to explore bootcamp participants' experiences and needs related to the bootcamp program. Specifically, our interview protocol included open-ended questions related to the bootcamp content, structure, teaching methods, and participants' interaction with their peers and the workshop facilitators. To analyze the qualitative data, we applied two cycles of coding. In Cycle 1, we applied a combination of open coding, in-vivo, and process coding methods [5,6]. The emerging codes from Cycle 1 helped us to create the initial codebook. In Cycle 2, we applied pattern coding to identify patterns both within and across participants' responses and organize them into larger themes and sub-themes [5,6]. The broader themes included reasons for enrolling into the program, program activities, interactions and collaborations, most enjoyable parts of the program, suggestions for improving the program, impact on future career decisions, and challenges.

3. Results and Conclusions

Overall, we found that the participants found the program beneficial and helpful for their careers. For example, some of the sub themes related to the most enjoyable parts of the program included relevant and up-to-date lecture topics, approachable lecturers, convenient timeline for busy professionals, making connections with instructors and peers, having a chance to make a difference for the country, among others. The participants found the program impactful in terms of their future career decisions. For instance, participants mentioned that they are planning to "design and deliver lectures on the topics covered at the bootcamp" and "apply the skills gained in clinical practice".

Our goal has been to train the leaders in a nascent field in an LMIC and our results indicate positive outcomes. More importantly, the training program has received tremendous support in the country, and we have successfully identified champions for the program.

References

[1] Shapiro LM, Kamal RN. Implementation of electronic health records during global outreach: A necessary next step in measuring and improving quality of care. The Journal of Hand Surgery. 2022 Mar 1;47(3):279-83. doi: 10.1016/j.jhsa.2021.09.016

[2] Mason C, Lazenby S, Stuhldreher R, Kimball M, Bartlein R. Lessons learned from implementing digital health tools to address COVID-19 in LMICs. Frontiers in Public Health. 2022;10. doi: 10.3389/fpubh.2022.859941.

[3] Chekijian S, Yedigaryan K, Bazarchyan A, Yaghjyan G, Sargsyan S. Continuing medical education and continuing professional development in the republic of Armenia: the evolution of legislative and regulatory frameworks post transition. Journal of European CME. 2021 Jan 1;10(1):1853338. doi: 10.1080/21614083.2020.1853338.

[4] Hersh WR, Bhupatiraju RT, Greene PS, Smothers V, Cohen C. Adopting e-learning standards in health care: competency-based learning in the medical informatics domain. In AMIA Annual Symposium Proceedings 2006 (Vol. 2006, p. 334). American Medical Informatics Association.

[5] Saldaña J. (2021). *The coding manual for qualitative researchers*. Sage Publications.

[6] Wicks D. The coding manual for qualitative researchers. Qualitative research in organizations and management: an international journal. 2017 Jun 12;12(2):169-70.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230179

Exploring Obese Adults' Preferences for a Physical Activity Chatbot: Qualitative Study

Dillys LARBI [a,b1], Rolf WYNN [b], Eirik ÅRSAND [a,c], Kerstin DENECKE [d], Paolo ZANABONI[a,b] and Elia GABARRON [a,e]
[a] *Norwegian Centre for E-Health Research, Tromsø, Norway*
[b] *Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø*
[c] *Department of Computer Science, UiT The Arctic University of Norway, Tromsø*
[d] *Institute for Medical Informatics, Bern University of Applied Sciences, Switzerland*
[e] *Department of Education, ICT and Learning Østfold University College, Halden*
ORCiD ID: Dillys Larbi https://orcid.org/0000-0002-1556-017X, Rolf Wynn https://orcid.org/0000-0002-2254-3343, Eirik Årsand https://orcid.org/0000-0002-9520-1408, Kerstin Denecke https://orcid.org/0000-0001-6691-396X, Paolo Zanaboni https://orcid.org/0000-0002-5469-092X, Elia Gabarron https://orcid.org/0000-0002-7188-550X

Abstract. Social media chatbots could help increase obese adults' physical activity behaviour. The study aims to explore obese adults' preferences for a physical activity chatbot. Individual- and focus group interviews will be conducted in 2023. Identified preferences will inform the development of a chatbot that motivates obese adults to increase their physical activity. The interview guide was tested in a pilot interview.

Keywords. Obesity, Physical activity, Social media, Chatbot, Qualitative research

1. Introduction

Regular physical activity is effective in managing obesity, which is estimated at about 23% among Norwegian adults [1]. Social media chatbots (i.e. chatbots integrated into a social media platform) could be an innovative way to increase regular physical activity in the adult population [2]. Our previous research confirms the importance of user involvement when developing a chatbot for physical activity [3]. Thus, we aim to explore obese adults' preferences for a social media chatbot for increasing their physical activity.

2. Methods

Individual and focus group interviews are scheduled for March 2023. Obese adults (≥18 years; body mass index > 30) will be recruited via posters and brochures distributed at

[1] Corresponding Author: Dillys Larbi, E-mail: dillys.larbi@uit.no, Tel: +47 909 49760.

Evjeklinikken, a specialist clinic for treating morbid obesity in Norway. The study protocol was declared exempt after assessment by the Norwegian Regional Ethics Committee (Ref: 351357). The University Hospital of North Norway's Data Protection Officer has approved the study (Ref: 2022/6610).

We will conduct 10 to 12 individual interviews using a semi-structured interview guide to explore obese adults' needs for features and functions of a physical activity social media chatbot. Two focus group interviews will be conducted with 10 to 12 participants each to discuss preferences for chatbot features and functions and receive feedback on the development of the chatbot using the paper prototype method. All interviews will be recorded, fully transcribed, anonymised, and analysed within an inductive content approach [4] using NVivo 12 Pro. In December 2022, we tested the interview guide in a pilot interview with a volunteer from Evjeklinikken.

3. Results and Discussion

The pilot interview with a 62-year-old female lasted for approximately 30 minutes. Reported preferences included: using a chatbot on either Facebook or Snapchat; adding physical activity challenges; using Norwegian language; clear and precise messages delivered 2-3 times per week; integrating a step counter; and a human chat option. Preliminary results of the interviews and focus groups will be available in April 2023.

The participant's preferences were based on recent experiences with social media platforms, activity-tracking devices, chatbots, and physical activity in general. Involving obese adults in the chatbot development process creates a sense of ownership and increases their interest and engagement with the resulting intervention [5]. Identifying obese adults' preferences for features and functions will inform the development of a chatbot that can motivate the users to increase their physical activity.

Acknowledgement

We thank Dr. Marianne V. Trondsen, the Evjeklinikken staff and the interviewee for their contribution. The study is funded by the Norwegian Centre for E-health Research.

References

[1] Bjørnelv GMW, Halsteinli V, et al. Modeling Obesity in Norway (The MOON Study): A Decision-Analytic Approach-Prevalence, Costs, and Years of Life Lost. Med Decis Making. 2021;41(1):21-36.
[2] Larbi D, Denecke K, et al. Usability Testing of a Social Media Chatbot for Increasing Physical Activity Behavior. J Pers Med. 2022;12(5).
[3] Larbi D, Sandsdalen H, et al. User preferences for a physical activity chatbot connected to an activity tracker and integrated into a social media platform. In: Henriksen A, Gabarron E, Vimarlund V, editors. Proceedings of the 18th Scandinavian Conference on Health Informatics; 2022 Aug 22-24; Tromsø, NO. Linköping (SE): LiU E-Press; c2022. p. 118-123
[4] Brinkmann S, Kvale S. InterViews: Learning the Craft of Qualitative Research Interviewing. Third ed. SAGE: Thousand Oaks; 2015.
[5] Gabarron E, Dorronzoro E, et al. Preferences and interests of diabetes social media users regarding a health-promotion intervention. Patient Prefer Adherence. 2018;12:2499-506.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230180

N-of-1 Analytics Makerspace: A Prototype for Cystic Fibrosis Self-Care and Co-Care

Andreas HAGER[a] and Carolina WANNHEDEN[b, 1]
[a]*Upstream Dream AB, Sweden*
[b]*Medical Management Centre, Department of LIME, Karolinska Institutet, Sweden*

Abstract. We have designed a prototype N-of-1 analytics makerspace, which is a collaborative work environment that provides a space for different stakeholders in healthcare to learn new skills and work together on projects that can improve individual patient care and the effectiveness of healthcare systems. Our prototype was designed to study the use of antibiotics in self-management for children with cystic fibrosis in Sweden but is intended to be disease agnostic and potentially include other complex medical conditions in the future.

Keywords. Patient-controlled real world data, N-of-1, knowledge commons, cystic fibrosis, learning networks.

1. Introduction

Different stakeholders – including healthcare managers, clinical researchers, patients, and care providers – have varying goals when it comes to understanding treatment variation [1]. Although different study designs may be needed to support analyses of treatment variation on an individual (N-of-1) or aggregated (population) level, they all build on data that are generated in individual patients' self-care and co-care (i.e., when patients' and other actors' resources are combined [2]). We have previously illustrated how using a knowledge commons framework can enable effective sharing of patient-controlled real-world data with different stakeholders, while respecting personal data privacy and control [3]. The aim of this study was to design an N-of-1 analytics makerspace that provides a hybrid environment for patient-controlled data management, governance, and analytics, allowing patients, providers, and researchers to collaborate in designing analytics to promote optimal treatment for individual patients and patient populations. Makerspaces are defined as "places where participants may work together to create and co-create knowledge and physical or digital products" [4].

2. Methods

Members of a collaborative learning network for cystic fibrosis (CF) care in Sweden developed a prototype N-of-1 analytics makerspace applied to CF self-care and co-care,

[1] Corresponding Author: Carolina Wannheden; Medical Management Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18a, Stockholm 171 77, Sweden; Email: carolina.wannheden@ki.se.

focusing on treatment variation in the use of antibiotics. The design process was based on community engagement: five adult CF patients and CF parents with special interest in quality improvement work contributed with their ideas and experiences.

3. Results

The prototype N-of-1 analytics makerspace includes three main components: 1) *N-of-1 analytics environment* that allows individuals to analyze and interpret data based on their own personal experiences or needs to gain insights and make informed decisions about their future self-care – this environment can also be used as a design space with de-identified or mock-up data; 2) *REDCap integration module* that enables smooth data sharing to support research projects; 3) *N-of-1 data governance platform* that contains a list of all datasets and variables with descriptions, supporting data quality conversations.

4. Discussion

Our prototype N-of-1 analytics makerspace illustrates how we can design a collaborative work and learning environment building on patient-controlled real-world data. Although the core data collection happens in the N-of-1 environment, data and insights can be shared with other stakeholder groups to support collaborative learning and knowledge development. In our continuing work, we will further develop the makerspace with user interfaces and export functionality for different stakeholder groups (e.g., care providers, clinical managers, clinical researchers). We will also test and evaluate the usability and usefulness of the prototype in a real setting.

5. Conclusions

By providing an N-of-1 analytics makerspace, we can support patient-centered collaboration among different stakeholders on projects that can improve individual patient care and the effectiveness of healthcare systems.

References

[1] Neuhauser D, Provost L, Bergman B. The meaning of variation to healthcare managers, clinical and health-services researchers, and individual patients. BMJ quality & safety. 2011 Apr 1;20(Suppl 1):i36-40. doi:10.1136/bmjqs.2010.046334.

[2] von Thiele Schwarz U. Co-care: Producing better health outcome through interactions between patients, care providers and information and communication technology. Health Services Management Research. 2016 Feb;29(1-2):10-5. doi:10.1177/0951484816637746.

[3] Hager A, Lindblad S, Brommels M, Salomonsson S, Wannheden C. Sharing patient-controlled real-world data through the application of the theory of commons: action research case study. Journal of medical Internet research. 2021 Jan 19;23(1):e16842. doi:10.2196/16842.

[4] Mersand S. The state of makerspace research: A review of the literature. TechTrends. 2021 Mar;65(2):174-86. doi:10.1007/S11528-020-00566-5/FIGURES/2.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230181

Medical Semiology Teaching Based on Intelligent eLearning

Eustache Muteba AYUMBA[1]
International Medical Informatics Association, Correspondent in DR Congo
ORCiD ID: Eustache Muteba A. https://orcid.org/0000-0002-2133-5710

Abstract. The paper proposes a methodology that emphasis techno-pedagogy, namely the constructivist and the adaptive intelligent learning of specialized semiology of COVID-19. An e-learning built on a constructivist pedagogy with a technology such adaptive intelligent environment, can be individualized (adaptive learning), can enhance learners' interactions with others (collaborative learning), and transforming the role of the teacher as facilitators of learning and assessor of competency. To make our system intelligent, we cope with Artificial Intelligence and Big data.

Keywords. Medical Semiology, Intelligent eLearning, COVID-19

1. Introduction

Since the advent of the COVID-19 pandemic, teaching is accompanied by technology which makes techno-pedagogy. There is a triangular relationship between teacher, machine and learner. From our point of view, constructivist learning and intelligent adaptive education are among the essential elements that allow to face the challenges posed by distance learning. The paper proposes an e-learning solution based on a constructivist and adaptive intelligent learning system of specialized semiology of COVID-19.

2. Methods

In accordance with WHO, primary care plays a significant role in gatekeeping and clinical responses: identifying and triaging possible COVID-19 cases, making an early diagnosis, helping vulnerable people cope with their anxiety about the virus, and reducing the demand for hospital services [1]. Semiology helps to build knowledge from the interpretation of subjective and objective data found on patient. Constructivist and adaptive intelligent learning can improve medical semiology teaching and the e-learning facilitates the acquisition of knowledge and skills [2]. An e-learning built on a constructivist pedagogy [3, 4] and in a technology such adaptive intelligent environment [5], can be individualized (adaptive learning), can enhance learners' interactions with

[1] Corresponding Author: Eustache Muteba Ayumba, P.O. Box 14769, Kinshasa, DR Congo, E-mail: emuteba@hotmail.fr.

others (collaborative learning), and transforming the role of the teacher as facilitators of learning and assessor of competency.

3. Results, Discussion and Conclusion

We present, in table 1, a methodology that emphasis the techno-pedagogy to support the development of an online medical education system. This is the first stage of our project.

Table 1. Constructivist and adaptive intelligent learning system.

Pedagogy	Techniques
Classroom: - Learner (Autonomous/Supported/Collaborative) - Teacher (Directed/Facilitated)	**Virtual Class**: Collaborative & Individual workspaces; Social interaction
Class description: Learning activities, tools, resources and materials	**E-valve & dashboard**
Content of knowledge/course: Pathophysiology, Clinical presentation, Pharmacotherapy, Case Repository & Ontology of COVID-19	**Electronic format**: texts, tables, graphics, images, sounds and videos.
Learner in learning: **-** Assimilation of knowledge; Consolidation of knowledge & Knowledge translation - Construction of knowledge and competencies based on ontology (pathway)	**Adaptive Intelligent learning environment:** Knowledge repository; profiling & intelligent tutoring + **Tools:** Organizers, Calendar and Contacts, Email, Videoconference & Instant Messaging
Evaluation: Problem-solving (individual & collaborative) based on interview; Physical Exam; Medical reasoning; Relevance of Additional & Presentation of case Exams	**E-assessment:** Individual interview, Report writing, Multiple choice question & Simulator of Clinical Cases + **Tools:** Evaluation-platform**,** Email & Videoconference
Teacher in mentoring: - Identification of the learner profile, adaptation of training according to performance, progressive profile refinement, recurring assessments & status of skills acquired - Analyze data coming from Analyzer & provide feedback directly or through the intelligent tutoring	**Analyzer:** based on profiler and e-assessment results + **Tools:** Organizers, Calendar and Contacts, Email, Videoconference & Instant Messaging

The scope of our study is to emphasis the techno-pedagogy, namely the constructivism and adaptive intelligent learning in medical education for medical students and health professionals seeking to strengthen their knowledge. We consider as stated in [6] that learning activities' role is a cornerstone to ensure the learners' learning process. The ontology plays also a major role in this process. In the other side, the teacher keeps the evolution of the learner on the dashboard based on information provided by the analyzer. To make our system intelligent, we cope with Artificial Intelligence and Big data.

References

[1] WHO. Role of primary care in the COVID-19 response. World Health Organization. 2021 Apr.
[2] Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med. 2006 Mar;81(3):207-12.
[3] Triantafyllou SA. Constructivist Learning Environments. Proceeding of the 5th International Conference on Advanced Research in Teaching and Education. 2022 Apr.
[4] Dennick R. Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. International Journal of Medical Education. 2016 Jun;7:200-205.
[5] Psyché V, Ruer P. L'apprentissage adaptatif intelligent. Pédagogie universitaire. 2019;8(4).
[6] Guerrero-Roldán AE, Rodríguez-González ME, Bañeres D, Elasri-Ejjaberi A, Cortadas P. Experiences in the use of an adaptive intelligent system to enhance online learners' performance: a case study in Economics and Business courses. Int J Educ Technol High Educ 2021;18:36.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230182

Support of Young Talent in the GMDS

Christina SCHÜTTLER[a,1] and Hannes ULRICH[b] on behalf of the GMDS Presidential Commission for the Support of Young Scientists
[a] *Central Biobank Erlangen, University Hospital Erlangen, Germany*
[b] *Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein, Germany*
ORCiD ID: Christina Schüttler https://orcid.org/0000-0001-9008-1868, Hannes Ulrich https://orcid.org/0000-0002-8349-6798

Abstract. The German Association for Medical Informatics, Biometry and Epidemiology (GMDS) aims to develop subject-specific methods, which are then to be applied in collaboration with various medical domains. Furthermore, the support of young scientists is an essential field of activity of the GMDS, since the need for junior staff has increased due to the acceleration in medical digitization. A specially established Presidential Commission strives to promote young talents and scientists in the above-mentioned disciplines. For this purpose, various strategies and concepts are elaborated in regular meetings and finally implemented. These include online formats such as a lecture series on research-related topics, as well as events such as summer schools and PhD symposia.

Keywords. Young scientists, lecture, summer schools, symposia, mentoring

1. Introduction

With currently about 2000 members, the GMDS[2] is the only scientific society that jointly represents the four disciplines of Medical Informatics, Biometry, Epidemiology and Medical Bioinformatics and aims to develop subject-specific methods that will then be applied in collaboration with various medical domains.

One of the mission statements is the support of young scientists of the GMDS. In recent years, the need for junior staff has increased due to the acceleration in medical digitization. To address this, the GMDS has established the Presidential Commission (PC) for the Support of Young Scientists, which seeks to inspire young talents, including the nearly 10% undergraduate GMDS members, within the above-mentioned disciplines.

2. Activities for the promotion of young professionals

The target group for which strategies are to be developed within the scope of the PC includes pupils, students, doctoral candidates, as well as postdocs. Accordingly, the following actions, among others, were developed and realized.

[1] Corresponding Author, Christina Schüttler, Central Biobank Erlangen, University Hospital Erlangen, Erlangen, Germany; E-Mail: christina.schuettler@uk-erlangen.de.

[2] German Association for Medical Informatics, Biometry and Epidemiology; https://www.gmds.de/de /

As general support GMDS offers young members low-threshold access to its community. In principle, young talents can become members of GMDS free of charge or reduced fee. In addition, the association provides financial support for students and young scientists to attend the annual meeting of the association. Outstanding theses are honored with sponsorship awards and young members with contributions at the annual meeting have the chance to win the Best Paper Award. As a point of contact, the PK operates a website[3] and a Twitter channel so that questions or suggestions can be addressed at any time.

A very successful format is the series of lectures offered on topics relevant to research. These take place online once a month during the semester and have a duration of 45 minutes plus 15 minutes of discussion. The content includes basic topics such as abstract writing and predatory journals, but also advanced topics such as clean coding or reproducibility of numerical experiments. In addition to lecturers from the GMDS community or from industry, foreign lecturers have also been recruited. The presentations, with a regular audience of 20 – 60 viewers, are also usually recorded and made available to GMDS members, establishing a sustainable knowledge base for current and subsequent members.

Face-to-face events take a particular place in networking. These include the doctoral symposia and the summer schools. The former is aimed at doctoral students in the first half of the doctoral phase. The event is composed of several parts. First, participants present their doctoral thesis, which they can then discuss with the other doctoral students. Another part serves for the acquisition of competences relevant to the doctoral thesis within the framework of workshops. Last but not least, social events are used to exchange experiences in a relaxed atmosphere.

The Summer Schools are on the other hand open to all interested parties, but are primarily aimed at students. Here, the focus is more on networking and integration into the GMDS community. With presentations from the respective disciplines, the fields of activity and possible areas of work are to be conveyed to the participants.

3. Outlook

A recent GMDS member survey [1] showed that the junior talent work is perceived, but could and should definitely be strengthened. Especially in the area of meetup formats. A special focus will be placed here when planning future activities. In addition, the PC seeks to connect with other young promotion groups on European level to provide a better promotion for junior scientists [2,3].

References

[1] GMDS member survey 2022. URL: https://www.gmds.de/fileadmin/user_upload/Publikationen/Mitteilungen/mitteil_gmds_2022_1.pdf

[2] Ognjanović I. yEFMI WG-Working Group of Young professionals and scientist within EFMI. Bilten Hrvatskog društva za medicinsku informatiku (Online). 2021 Jul 11;27(2):45-6.

[3] Gallos PG, Mantas J, Pape-Haugaard L, Randell R, Stoicu-Tivadar L, Weber P, Winter A. EFMI. Yearbook of Medical Informatics. 2021 Aug;30(01):340-5.

[3] Presidential Commission Website: https://www.gmds.de/de/ueber-uns/organisation/praesidiumskommissionen/nachwuchsfoerderung-in-der-gmds/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230183

Reduction in X-ray Retake Rate Using the Token Economy Method

Hiroto HADA[a,b,1] , Shusuke HIRAGI[a, c] , Yukiko MORI[a] , Kenichi SAITO[a] , Masumi AI [b] and Tomohiro KURODA [a]
[a] *Kyoto University Hospital, Kyoto, Japan*
[b] *Tokyo Medical and Dental University Hospital, Tokyo, Japan*
[c] *PIIF Tazuke-kofukai Medical Research Institute KITANO HOSPITAL, Osaka, Japan*
ORCiD ID: Hiroto HADA https://orcid.org/0000-0003-1192-2839

Abstract. Excessive retakes of X-ray images increase labor and material costs, as well as result in excess radiation exposure for patients and a long waiting time. In this study, we evaluated the effectiveness of the token economy method as a management method for reducing X-ray retake rate among radiology technicians. The results showed a 2.5% reduction in retake rate, indicating the effectiveness of our method. In addition, we suggest that the token-economy-based approach can be applied to other hospital management problems.

Keywords. Hospital Management, Token Economy, X-ray retakes

1. Introduction

Radiology technicians sometimes fail to obtain satisfactory radiographs; therefore, retakes are unavoidable. However, excessive retakes increase labor and material costs, as well as result in excess radiation exposure for patients and a long waiting time. Hence, department managers should reduce the retake rate while maintaining image quality. In this study, we implemented and evaluated the token economy method [1] to modify technicians' behavior so that more attention is paid to obtaining a satisfactory image at once.

2. Method

First, we designed a token economy to achieve the aforementioned goals. Specifically, we distributed 200 units of token at the beginning of the intervention, and technicians received 11 units of tokens when they successfully submitted the radiographic image, whereas they had to pay 10 units when they took them. We set these units based on the target retake rate of 10%. Next, to evaluate our method, we compared the retake rates of the two periods before implementing the token economy (1st phase: no specific intervention, 2nd phase: individual retake rates were pinned up for mutual comparison), and periods after implementing the token economy. In this study, we limited the target to skeleton radiography imaging, which requires technicians' skill; therefore, retake rates

[1] Corresponding Author: Hiroto HADA, E-mail: hada.hiroto.l37@kyoto-u.jp.

should be manipulated with their effort. We recruited seven technicians from a university hospital in Japan and calculated the retake rates among these phases. This study was approved by the Kyoto University Hospital Ethics Committee (approval number: R3525)

3. Results

The retake rates for the study participants were 10.7 ± 3.6% (mean ± standard deviation [SD]), 8.1 ± 1.7%, 5.6 ± 1.4% at the no-intervention, pinned-up, and token economy phases, respectively. Changes in the individual retake rates are presented in Table 1. Participants eventually accumulated 530–1140 units after 91 working days in the intervention period. To describe the accumulation process, for example, a participant initially loses 105 units of token but the participant's retake rate subsequently declines and they finally accumulate 909 units of token.

Table 1. Average retake rate for each period

Staff	A	B	C	D	E	F	G	Mean (SD)
No intervention	8.1	13.6	10.7	6.9	12.8	16.1	7.0	10.7 (3.6)
Pinned retake rates	6.6	9.1	10.3	5.7	8.7	9.5	6.8	8.1 (1.7)
Token economy	5.8	5.9	7.0	2.7	6.9	5.8	5.0	5.6 (1.4)

4. Discussion

Visualizing the number of available resources and self-management of consumed resources are hypothesized to lead to behavioral change among healthcare professionals [2]. Token economy is a visualization method based on a learning theory called operant conditioning. This behavioral psychological method transforms extrinsic motivation into intrinsic motivation [3]. As an application of behaviorist psychology, the current study was conducted to test management techniques in a clinical setting, based on changes in retake rates and token circulation. The results showed that the retake rate decreased for all staff members after the token economy intervention; on average, a 2.5% reduction in retake rate was observed in all staff in the token economy period compared to the "pinned" period. In addition, some participants showed improved retake rate after checking token reduction in their "wallets". The retake rate for all staff members fell below the organization's target of 10%. These results indicate that the effectiveness of the proposed method at reducing the retake rate. In addition, we suggest that the token-economy-based approach can be applied to other hospital management problems.

References

[1] Hiragi S, Hatanaka J, Sugiyama O, Saito K, Nambu M, Kuroda T. Token Economy–Based Hospital Bed Allocation to Mitigate Information Asymmetry: Proof-of-Concept Study Through Simulation Implementation. JMIR Form Res. 2022 Mar 4;6(3): e28877, doi: 10.2196/28877.

[2] Gross R, Nirel N, Boussidan S, Zmora I, Elhayany A, Regev S. The influence of budget-holding on cost containment and work procedures in primary care clinics. Soc Sci Med. 1996 Jul;43(2):173-86, doi: 10.1016/0277-9536(95)00359-2.

[3] Boniecki KA, Moore S. Breaking the silence: using a token economy to reinforce classroom participation. Teach Psychol. 2016 Aug 26;30(3):224-227, doi: 10.1207/s15328023top3003_05

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230184

Fairness in Artificial Intelligence: Regulatory Sanbox Evaluation of Bias Prevention for ECG Classification

Arian RANJBAR[a,1], Kristin SKOLT[b], Kathinka Theodore AAKENES VIK[b], Beate SLETVOLD ØISTAD[b], Eilin WERMUNDSEN MORK[b] and Jesper RAVN[b]
[a]*Medical Technology and E-health, Akershus University Hospital, Norway*
[b]*Norwegian Data Protection Authority, Oslo, Norway*
[c]*Equality and Anti-Discrimination Ombud (LDO), Oslo, Norway*

Abstract. As the use of artificial intelligence within healthcare is on the rise, an increased attention has been directed towards ethical considerations. Defining fairness in machine learning is a well explored topic with an extensive literature. However, such definitions often rely on the existence of metrics on the input data and well-defined outcome measurements, while regulatory definitions use general terminology. This work aims to study fairness within AI, particularly bringing regulation and theoretical knowledge closer. The study is done via a regulatory sandbox implemented on a healthcare case, specifically ECG classification.

Keywords. Artificial Intelligence, Fairness, Bias, Ethics, Regulation, GDPR

1. Introduction

As Artificial intelligence (AI)-enabled solutions within healthcare are getting closer to production, a raised attention has been directed towards potential ethical issues. This development includes regulation of AI which increasingly use terminology or requirements of fairness and fair use.

Defining fairness has a long history within ethics. Although extensive research has been conducted on the topic, there is no general consensus on a definition. On the other hand, to provide fair models within machine learning, fairness needs to be quantifiable. Such algorithmic fairness definitions have previously been widely explored, [1]. Analysis often relies on either individual fairness, i.e. "mapping similar people similarly", or statistical parity. Common among them is the need to set the measurement with respect to a metric, protected attributes and desired outcomes.

This work aims to investigate fair use of AI within healthcare from a regulatory perspective. How is fairness defined, and what can be done to ensure fairness? More importantly, can algorithmic methodology and regulatory frameworks be aligned? The investigation is carried out within a regulatory sandbox on a case study of an AI-enabled ECG classifier.

[1] Corresponding Author: Arian Ranjbar, Akershus Universitetssykehus HF, 1478 Lørenskog, Norway; E-mail: arian.ranjbar@ahus.no. This research was partly funded by Nasjonalforeningen for folkehelsen.

2. Method

To evaluate the fairness and bias definitions from a regulatory perspective, a regulatory sandbox is implemented together with the corresponding authorities. The sandbox as a method was developed to enable testing of new technology or methodology, currently not (or recently) covered by or compliant with existing regulatory frameworks, [2].

3. Case Study

The case study regards a machine learning based ECG decision support system, developed internally at Akershus University Hospital using regular production systems, [3]. Relevant regulation is found in GDPR and Norwegian Anti-Discrimination Act.

A very brief summary of the outcomes follows. Fairness is central to several regulations such as GDPR. However, it has no clear definition, but rather follows a dynamic principle, i.e. adjusts over time in accordance to general societal perception. Corresponding guidelines also keep advice at a high level, not generally compatible with explicit quantification. The Norwegian equality and Anti-Discrimination Act, elaborate on fairness a a prohibition of discrimination with respect to twelve specified protected grounds, e.g. gender, ethnicity, age etc.; useful from a quantitative perspective. Bias similarly has no legal definition, however is often referred to in relation to fairness. In this regard four sources of bias are specifically pointed out: bias arising from already existing biases within healthcare, data collection, design- and deployment principles, and application injustice or misuse; all extensively researched within machine learning, [1].

Although the particular risk of bias in the ECG algorithm is judged as low, countermeasures may be necessary from a regulatory perspective such as: data control, monitoring systems with re-training capabilities, and establishment of training routines for clinicians. Notably, from a regulatory perspective, the full treatment chain may be considered in a fairness evaluation. Thus, even if a specific algorithm produces biases, it might be considered fair if patients receive the same treatment in the end.

4. Conclusion

This study took the first step in aligning fairness from a regulatory and algorithmic perspective, using the sandbox methodology. Particularly finding explicit traits to protect for and necessary bias reduction procedures. However, evaluation may still be challenging since information regarding protective grounds is generally not available, and may be in conflict with other regulation. Further research is also needed regarding metrics and outcomes to optimize for, with respect to compliance.

References

[1] Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR). 2021;54(6):1-35.

[2] Leckenby E, Dawoud D, Bouvy J, Jonsson P. The sandbox approach and its potential for use in health technology assessment: a literature review. Applied health economics and health policy. 2021.

[3] Ranjbar A, Ravn J, Ronningen E, Hanseth O. Enabling Clinical Trials of Artificial Intelligence: Infrastructure for Heart Failure Predictions. Proceedings of Medical Informatics Europe. 2023.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230185

Quantitative Study of a Regional Patient Portal Usage in the Pandemic Period

Javier QUILES[a,1], Ricardo LÓPEZ [b] and Benigno ROSÓN[b]
[a] *Servizo Galego de Saúde, Area Sanitaria de Ferrol, Servicio de Informatica, Spain*
[b] *Consellería de Sanidade, Sistemas y Tecnologías de la Información, Spain*
ORCiD ID: Javier Quiles https://orcid.org/0000-0001-5945-0178; Ricardo López https://orcid.org/0000-0001-5416-7066; Benigno Rosón: https://orcid.org/0000-0002-6308-1480

Abstract. In 2013 using a Public Procurement of Innovation procedure the Region of Galicia developed a patient portal called "E-Saúde", that went live in 2015. COVID situation in 2019 produced a high demand of e-health services, scaling by 10x the number of users in 2021. *Objective:* In this study a quantitative description of patient portal usage from 2018 to 2022 is made to show the behaviour of usage trends of a patient portal before, during and after COVID pandemic. *Methods:* Two main data sets were obtained from patient portal logs to obtain: 1) Enrolment of new users and number of sessions opened in the portal. 2) Detailed usage of relevant functionalities. Descriptive statistical methods were applied to show the usage of the portal in a biannual time series description. *Results:* Prior to the pandemic, the portal was gradually being introduced to citizens. During pandemics, more than 1 million users were registered and a peak of 15x usage could be observed. After COVID, the level of usage of portal services decreased, but kept a sustained trend five times higher than in Pre-COVID situation. *Conclusion:* There is limited information available on metrics, functionalities and acceptability for general purpose patient portals, but the analysis performed on usage levels shows that after a high peak reached during COVID period, explained by the need of direct access to clinical information, the level of usage of the patient portal remains five times higher than in pre-pandemic situation for all functionalities of the patient portal.

Keywords. Patient Portal, COVID, Information Technology, eHealth, EHR.

1. Introduction

Seven years after going live, the objective of the study is to show the behaviour and level of usage of the Galician regional patient portal "E-Saúde" before, during and after COVID pandemic, and provide information of usage towards a "new normal" situation [1].

2. Methods

Log data from frontend and backend servers was used to obtain time series of data values of usage corresponding to each month from 1 January 2018 to 31 December 2022.

[1] Corresponding Author: Javier Quiles del Río, E-mail: jquirio@sergas.es.

Two main data sets were extracted, aggregated and represented to show 1) the biannual average number of new users and number of accesses in the portal - Figure 1-, and 2) the detailed usage of the most relevant functionalities -Figure 2-.

SEMESTER-YEAR	NEW USERS	TOTAL ACCESS
1-2018	16.158	246.127
2-2018	13.956	277.721
1-2019	18.763	423.973
2-2019	18.863	471.949
1-2020	16.586	560.383
2-2020	25.635	831.149
1-2021	250.244	2.051.660
2-2021	564.633	6.132.926
1-2022	83.269	4.090.802
2-2022	59.114	3.264.411

Figure 1. Graph and Number of users and unique accesses by semester

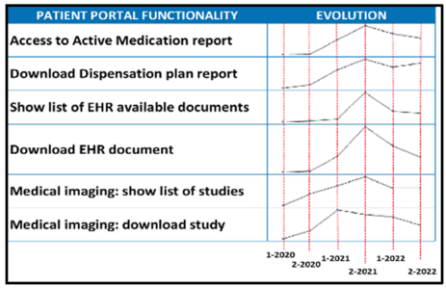


Figure 2. Graph showing usage of relevant functionalities by semester

3. Results

In 2018, prior to the pandemic, the portal had been used by more than 100.000 citizens. During pandemics, more than 1 million users gained access to the portal. A peak of 15x in all measured elements can be observed. After COVID, the level of usage decreased, but shows a sustained usage level at least five times higher than in Pre-COVID situation.

4. Discussion

COVID-19 pandemic period has brought significant changes on healthcare, that will not come back to its previous situation. The high level of usage that was observed is explained by the need of direct access to clinical information. This increased level of usage has become the baseline for the next "new-normal" period, after a critical mass of citizens have gained access to their clinical information [2].

5. Conclusion

At the end of pandemic period, from june 2022 onwards, values of patient portal usage show a sustained trend that is five times higher than in preCovid period. Further development of patient portals are expected to bring a positive effect on healthcare transformation, patient autonomy and empowerment as a natural solution.

References

[1] Morrison, Ciarán, Rimpiläinen, Sanna, Bosnic, Iris. Emerging Trends in Digital Health and Care : A Refresh Post-COVID [Internet]. Digital Health & Care Institute; 2022 [cited 2022 Dec 10]. Available at: https://strathprints.strath.ac.uk/id/eprint/82203

[2] Ammenwerth E. From eHealth to ePatient: The Role of Patient Portals in Fostering Patient Empowerment. ejbi [Internet]. 2018 [cited 2022 Dec 8];14(2). Available from: https://www.ejbi.org/scholarly-articles/from-ehealth-to-epatient-the-role-of-patient-portals-in-fostering-patient-empowerment.pdf

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230186

Mobile Application for Improvement of Self-Management of Type 2 Diabetes: Usability Pilot Test

Andreia PINTO [a,1], João VIANA [a], Glória CONCEIÇÃO [a], Paulo SANTOS [a], Cristina SANTOS [a] and Alberto FREITAS [a]
[a] *CINTESIS@RISE, MEDCIDS, Faculty of Medicine of the University of Porto, Portugal*

Abstract. We intend to evaluate the usability of a mobile app developed for the self-management of T2DM. A pilot usability cross-sectional study was performed with a convenience sample of 6 smartphone users aged 45 years. Participants performed tasks autonomously in a mobile app to assess if users could complete them and filled out a usability and satisfaction questionnaire. About half of the tasks had a successful completion rate. The result of the usability questionnaire was 64/100, below the acceptable value, but the satisfaction value was considered good. This study was fundamental as it allowed us to verify which improvements should be implemented in the next version of the app, contributing to its better acceptance.

Keywords. Usability, tasks, mobile applications, Type 2 diabetes mellitus, T2DM

1. Introduction

Diabetes is one of the most prevalent diseases worldwide [1]. Taking into consideration the growing use of mHealth, the number of apps aiming to help patients with type 2 diabetes (T2DM) patients has increased. However, not all of them truly test whether the app meets the needs of its target audience [2]. This paper aims to evaluate the usability of the first version of a mobile app developed to help manage T2DM.

2. Methods

A pilot usability cross-sectional study was performed. We recruited a convenience sample of non-diabetics who were smartphone users and aged 45 years or older, as we just wanted to verify if the application had good usability. The mobile application has features to improve the lifestyle behavior of the participants, in particular for controlling and monitoring food, physical activity, medication, and glycemic control. Participants were asked to perform a set of 13 pre-defined tasks created to evaluate the implemented features, in particular, to check the operation of the food detection feature. In addition, the participants completed a usability questionnaire, namely the System Usability Scale (SUS), and also a satisfaction scale: User Satisfaction Evaluation Questionnaire (USEQ). The Ethics Committee Faculty of Medicine of the University of Porto approved the study.

[1] Corresponding Author: Andreia Pinto, E-mail: andreiapinhopinto@gmail.com

3. Results

The sample was composed of six participants, of which 67% were female and with an average age of 59 years-old. A maximum percentage of success was observed in about half of the tasks. The tasks that were least successful were mostly those that asked to edit what had been previously added. All participants were successfully able to identify a food through the photography option, though the meal recognition system was not always accurate. Regarding the questionnaires completed by the participants, a score of 64/100 was obtained according to the SUS, which was considered below the acceptable value (<68). On the other hand, considering a maximum of 30 possible points, the USEQ scale obtained a value of 23.5, which was considered that the app had good satisfaction.

4. Discussion

The pilot usability study was conducted to verify if the first version of our mobile app had good usability before we implement the next features. We verified that the food recognition system by photography was still not adequately accurate, and will need further improvements in next versions. However, participants understood how to use this functionality, since it was successfully completed by all of them. The resulting value of the score obtained by the SUS follows the identified problems, thus, being below what is considered acceptable for good usability, indicating that the app needs improvements. However, despite this, satisfaction was considered good and this is one of the essential steps for the intention of continuing to use the app and reducing the dropout values.

5. Conclusions

Despite having carried out a small study and with a convenience sample, it was possible to learn lessons for the next version and to obtain an app suited to the needs of users.

Acknowledgment

This research work was developed under the project Food Friend - “Autonomous and easy-to-use tool for monitoring of personal food intake and personalized feedback (ITEA 18032), co-financed by the North Regional Operational Program (NORTE 2020) under the Portugal 2020 and European Regional Development Fund (ERDF) with the reference NORTE-01-0247-FEDER-047381. Andreia Pinto was funded by FCT, Portugal under PhD grant number UI/BD/151485/2021.

References

[1] Whiting DR, Guariguata L, Weil C, Shaw J. IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Research and Clinical Practice. 2011;94(3):311-21.

[2] Baptista S, Trawley S, Pouwer F, Oldenburg B, Wadley G, Speight J. What Do Adults with Type 2 Diabetes Want from the "perfect" App? Results from the Second Diabetes MILES: Australia (MILES-2) Study. Diabetes Technology and Therapeutics. 2019;21(7):393-9.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230187

The Demonstration of a Tool for Self-Estimating Digital Competence

Karin AHLIN[1]
Service Research centre, Karlstad University, Sweden
ORCiD ID: Karin Ahlin https://orcid.org/0000-0003-4051-6960.

Abstract. This study presents the results from a demonstration of a tool for self-estimation of digital competence for nurses and assistant nurses. The data was gathered from twelve participants working as leaders of older care homes. The results show that digital competence is of importance in health and social care, that the dimension of motivation is of utmost importance and that the presentation of the survey results should be flexible.

Keywords. Digital competence, health and social care, tool, demonstration

1. Introduction

An increasingly large part of public organizations' work tasks requires digital knowledge, such as nurses signing a patient's medical file while giving them their prescriptions. Because digital competence is fundamental for working [1], it should be understood and developed with the co-workers. One part of the development is to understand the status of a co-worker's digital competence and how to develop it. Therefore, the aim of this study is to demonstrate a tool for self-estimation of co-workers' digital competence in public organizations.

Schiefloe [2] emphasizes digital transformation as organizational changes initiated and shaped when digital technology is introduced and builds the Pentagon model on five dimensions: (1) formal structure, (2) technology and infrastructure, (3) social relations and networks, (4) interaction and (5) culture and competence. The *formal structure* describes how responsibility and accountability within an organization are distributed based on departments and work roles. *Technology and infrastructure* refer to the equipment the organization's members use to perform their tasks. *Culture and competence* can be summarized as the values, language, attitudes, working methods and ways of making decisions in the organization. *Interaction* means how the organization's employees relate to each other, especially in work processes, to achieve common goals. Interaction can be divided into cooperation/collaboration, communication, coordination and management. *Social relationships and networks* are divided into informal relationships that exist internally and externally and are used to optimize the organization's results. Added to the Pentagon model's dimensions is motivation, which is viewed as the individual's ability and willingness to change [3].

[1] Corresponding Author: Karin Ahlin, E-mail: karin.ahlin@kau.se.

2. Methods

The method used here is a part of Design Science Research (DSR), namely the demonstration step. The case for the study is the VälTel2.0 project [1]; a project including cities and the region in Mid Sweden and Tröndelag in Norway. The demonstration was done via two half-day workshops and had discussed the use of digital tools, the view of the dimensions, and the survey results presentation. The twelve participants all held a leading function in various elder care homes. The workshops were recorded and transcribed, and analyzed according to the described themes.

3. Results and Conclusions

All respondents declared that digital tools in health and social care are used for everyday work assignments like digital signing for medications, record keeping, scheduling, and email and therefore digital competence is of high interest for the organization. For the *dimension Formal Structure* the respondents emphasized that the internal organization must prepare and adapt work processes and guidance to be applicable to their specific user group. In the *dimension Technology and Infrastructure* was the user interface and the number of systems employees needed to use; respondents described how these affected the efficiency of their work. The respondents emphasized the *dimension Culture and Competence* as important and viewed the organization's culture as significant for influencing ongoing learning about digital tools. Knowledge exchange about digital tools often took place directly between employees, for example, during work or coffee breaks. The *dimension Interaction* in the form of communication and information was viewed as an essential part of the organization. One sign that internal interaction is essential is that one is aware of communication preferences, for example, via Teams or Outlook. *The dimension Social Relations and Networks* was viewed as less important in healthcare since the demand for building social relationships and networks with the help of digital tools is not necessary. Therefore, this dimension can be removed while self-estimating digital competence. *The dimension Motivation* was viewed as essential while developing digital competence. It is seen as so necessary that those with high motivation are often asked if they want to be digital ambassadors and participate in, for example, implementation. The presentation of the results from the survey must be flexible and adaptable to the recipient group. Visualizations and descriptive statistics, such as spider diagrams, traffic lights, and mean value, are essential and should be easy to use. Based on the results, the respondents highlighted the importance of targeted training resources to purchase training. Synthesizing the results show that the dimensions were all viewed as essential, besides the dimension of Social relations and networks, due to its irrelevance to the health and social care context. The self-estimation of the individual's motivation was judged to be of utmost importance during the evaluation, even so vital that it could mean new tasks for the co-workers.

References

[1] Ahlin K, Langlo JA. Rapport Digital Mognad–gemensamma resultat VälTel 2.0. 2022.
[2] Schiefloe PM. Mennesker og samfunn: innføring i sosiologisk forståelse. 2003.
[3] Heggerners TA. Digital forretningsforståelse: Vigmostad & Bjørke; 2013.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230188

E-learning as Part of Residency Education

Adam ABOVARDA[a], Helena VALLO HULT[a,b,1], Christian MASTER ÖSTLUND[a] and Paul PÅLSSON[b,c]
[a] *University West, School of Business, Economics and IT, Sweden*
[b] *NU Hospital Group, Sweden*
[c] *University of Gothenburg, Sweden*

Abstract. Through a literature review in combination with qualitative analysis of course evaluations, this study examines aspects that contribute to enhancing e-learning for physicians in a residency education program. The literature review and the qualitative analysis outline three main factors (pedagogical, technological, and organizational), highlighting the importance of a holistic approach that includes learning and technology in context when integrating e-learning strategies in adult learning programs. The findings contribute insights and practical guidance for education organizers on how to conduct e-learning during and after the pandemic.

Keywords. Covid-19, continuing professional development (CPD), residency education, digitalization, e-learning, healthcare, physicians.

1. Introduction

Like many education institutes, organizers of medical residency education had to switch to e-learning during the COVID-19 pandemic. In healthcare e-learning, individual digital skills and understanding of the usefulness of digitalization in practice, along with pedagogical factors, self-motivation and interactive learning methods can be facilitating factors [1,2]. This study contributes to enhancing e-learning for resident physicians in Sweden through the following research question: *how do resident physicians perceive and relate to e-learning, and what aspects should be considered when integrating e-learning as part of adult learning?* The paper adopts a sociotechnical approach to e-learning, stressing the importance of social situations and interactions, and technology aspects to foster an environment that creates engagement and motivation [3,4].

2. Methodology

The study was performed by conducting a literature review [5] combined with qualitative analysis on course evaluations, to investigate physicians' experiences, perceptions, and satisfaction with e-learning. Four online courses with resident physicians (79 individuals, 119 participants in total) in a Swedish hospital during 2020-2021 formed the base for the analysis. The questionnaires contained open-ended reflective questions about participants' experiences and were analyzed through thematic analysis [6].

[1]Corresponding Author: Helena VALLO HULT, University West and NU Hospital Group, Trollhattan, Sweden; E-mail: helena.vallo-hult@hv.se.

3. Results

The main factors influential for e-learning identified in the literature review can be summarized as pedagogical (e.g., learner-instructor interaction), organizational (e.g., infrastructure, training) and technological (e.g., hardware and software). The qualitative analysis correspondingly revealed three main themes related to the physicians' perceptions, experiences and satisfaction with e-learning: i) views and ideas related to the technology used during their participation in the online courses. This could be about internet connection or sound and video using Zoom. ii) pedagogy-related issues, which could be about the lectures, group discussions and social interactions; and iii) course management issues, such as planning and disseminating of course materials.

4. Discussion and Conclusion

The participants in the study valued the enabling *pedagogical factors* of e-learning, such as facilitated discussions in break-out rooms and walk-and-talk groups. This is in line with the literature, where interactive functionalities were considered among the most important features [4,7]. *Technology factors,* such as reliable tools and connection, are among the characteristics that affect the usability and the effectiveness of e-learning [8]. Participants in this study were likewise unsatisfied when faced with technology-related issues such as poor audio quality and the absence of webcams by some of their colleagues. *Organization factors* related to management of course planning and activities, were also mentioned, such as access to course material, and communication during lectures [7]. In sum, similarities in the literature review and the qualitative analysis confirm that all factors should be considered, highlighting the importance of a socio-technical approach that includes learning and technology in context [3,4]. Altogether, the findings suggest that important lessons are to be learned, but in doing so, there is great potential for integrating e-learning strategies into post-pandemic residency education.

References

[1] Tsodikova O, Korzh O, Hyria M. Information and communication technologies in postgraduate training of primary care doctors: a new look at the problem of using online resources during the Covid-19 pandemic. Technium Soc Sci J. 2020;8:60.
[2] Vallo Hult H, Hansson A, Gellerstedt M. Digitalization and Physician Learning: Individual Practice, Organizational Context, and Social Norm. Journal of Continuing Education in the Health Professions. 2020;40(4):220-7
[3] Upadhyaya KT, Mallik D. E-Learning as a Socio-Technical System: An Insight into Factors Influencing its Effectiveness. Business Perspectives and Research. 2013;2(1):1-12.
[4] Vallo Hult H, Islind AS, Master Östlund C, Holmgren D, Wekell P. Sociotechnical Co-design with General Pediatricians: Ripple Effects through Collaboration in Action. AMCIS, USA 2020.
[5] Ramdhani A, Ramdhani MA, Amin AS. Writing a Literature Review Research Paper: A step-by-step approach. International Journal of Basic and Applied Science. 2014;3(1):47-56.
[6] Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
[7] Ouadoud M, Rida N, Chafiq T. Overview of E-learning Platforms for Teaching and Learning. Int J Recent Contributions Eng Sci IT. 2021;9(1):50-70.
[8] Al-Fraihat D, Joy M, Sinclair J. Evaluating E-learning systems success: An empirical study. Computers in human behavior. 2020;102:67-86.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230189

Proposing a Novel Hybrid Short-Term Exchange Program in Biomedical and Health Informatics Education

Thomas ENGELSMA[a,1], Carolin ANDERS[b], Susan J. OUDBIER[c], Karen EILBECK[d], Petra KNAUP[b], Linda W. PEUTE[a] and Matthias GANZINGER[b]
[a] *Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Amsterdam Public Health research institute, Amsterdam, The Netherlands*
[b] *Heidelberg University Hospital, Institute of Medical Informatics, Heidelberg, Germany*
[c] *Amsterdam UMC location Vrije Universiteit Amsterdam, Outpatient Division, Amsterdam Public Health research institute, Amsterdam, The Netherlands*
[d] *Department of Biomedical Informatics, School of Medicine, University of Utah, Salt Lake City, USA*

Abstract. International student exchange is a valuable opportunity for Biomedical and Health Informatics students to gain new perspectives and experiences. In the past, such exchanges have been made possible through international partnerships between universities. Unfortunately, numerous obstacles such as housing, financial concerns, and environmental implications related to travel, have made it difficult to continue international exchange. Experiences with hybrid and online education during covid-19 paved the way for a new approach that allows for short international exchange with a hybrid online-offline supervision model. This will be initiated with an exploration project between two international universities , each related to their respective institute's research focus.

Keywords. Medical Informatics, Education, International Exchange, IPHIE

1. Introduction

International exchange provides students with experiences that propel their future career. Six university programs in Biomedical and Health Informatics (BMHI) education from various international institutes across Europe, Asia, and the United States have been cooperating for several decades. This cooperation known as the International Partnership in Health Informatics Education, or IPHIE, goes back to 1998 [1]. It aims to prepare faculty and students in BMHI for top international positions in medical information and communication technology with an overarching objective to establish an international network for training and education in this field where they can learn from each other. IPHIE organizes bi-annual masterclasses, shares curricula, and supports faculty and student exchanges for a longer period of time in line with IMIA educational objectives for BMHI [2]. However, student housing and the availability of finances and other

[1] Corresponding Author: Thomas ENGELSMA, E-mail: t.engelsma@amsterdamumc.nl.

resources for such travel-intensive activities all present difficulties. Also, new challenges arise related to the environment and sustainable solutions for these travel-intensive activities. These challenges ask for novel, perhaps experimental, ways to support international student exchanges that include less travel-activities and resources, but still provide a valuable international experience for BMHI students. In this paper, we propose a new initiative for a short-term international exchange between two universities, based on our experiences with hybrid and online teaching during the COVID-19 pandemic. This initiative aims to enhance international collaboration, support BMHI research and share best-practices between international universities.

2. Proposed Exchange Program

The proposed exchange program initially involves two universities involved in IPHIE, Heidelberg University hospital (UKHD) in Germany and the Amsterdam UMC (location University of Amsterdam) in the Netherlands. Both institutes are conducting research in the context of mobile Health and are experiencing similar challenges. The Amsterdam UMC performs research on developing inclusive patient satisfaction questionnaires of digital health tools and UKHD is developing a tablet computer based system for digital anamnesis. In both institutions it was found that patients can experience problems understanding the questions because they do not comply with the average language levels of B1. This topic was introduced in assignments during courses on eHealth literacy in both universities' Medical Informatics program. Afterwards, a proposal for a joint four month internship was distributed among students at each site. Supervision for these internships will be provided jointly by both institutions. The students will visit the other university for two to four weeks to have the opportunity to learn from faculty and other students and experience international exchange. Challenges related to housing and finances for international exchange are attenuated with this shorter exchange. After the exchange, a systematic evaluation will be conducted by interviewing the student(s) to assess their experiences and satisfaction with the program, as well as their perceived academic and personal benefits.

This paper proposed a new initiative for short-term international research projects between two universities to facilitate student exchange and medical informatics research. The exchange program will encourage collaboration between the two or more universities, allow the sharing of best practices and research methods between both sites, and ease the opportunities for students to participate in an international research project. We believe that this exchange program will offer an unique opportunity for students to gain experience in medical informatics research and learn from one another from a societal, cultural, and professional perspective. If the evaluation of this approach is positive, we aim to expand this with other institutes involved in IPHIE and encourage other universities to implement this approach as well.

References

[1] Jaspers MW, Gardner RM, et al. IPHIE: An International partnership in health informatics education. InMedical Infobahn for Europe 2000 (pp. 549-553). IOS Press.

[2] Bichel-Findlay J, Koch S, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics: Second Revision. International Journal of Medical Informatics. 2023 Feb 1;170:104908.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230190

Expert Feedback on the Adaptation and Translation of Spanish Version of WiseApp

Rebecca SCHNALL[a,1] Sergio Ozoria RAMIREZ,[a] Janeth Juarez PADILLA,[a]
Mina HALPERN[b] Susan OLENDER[a] and Pamela BAEZ[b]
[a] *Columbia University Medical Center, NY, NY, USA*
[b] *Clinica de Familia, La Romana, Dominican Republic*

Abstract. HIV-related disparities also exist in developing countries, such as the Dominican Republic, where minority groups and/or those with low socioeconomic status experience higher disease burdens and worse health outcomes than those with higher socioeconomic status. We used a community-based approach to ensure the WiseApp intervention is culturally relevant and addresses the needs of our target population. Expert panelists made recommendations on how to simplify the language and features of the WiseApp to accommodate Spanish-speaking users who may have lower levels of education, or color or vision deficiencies.

Keywords. mHealth, HIV, Spanish, usability, community-based participatory research

1. Introduction

Marked deficits remain in antiretroviral therapy (ART) adherence and viral suppression among persons living with HIV (PLWH) particularly among Spanish-speaking Latinos in the US and the Caribbean [1]. To that end, our team conducted extensive formative work with PLWH and their providers to identify the content, features, functionalities, and interface of a mobile app (WiseApp) to improve health outcomes, including ART adherence. Building on the findings, the Agency for Healthcare Research and Quality provided support to build this app and integrate it with a smart pill box, enabling PLWH to self-monitor their medication adherence in real-time. The WiseApp supersedes current approaches to ART adherence using a self-management mobile app linked to an electronic pill box and was found to be efficacious for improving medication adherence in a randomized controlled trial (NCT03205982) [3]. At the same time, HIV disproportionately affects Latinos who experience later diagnoses and face more barriers to testing and treatment than their White counterparts [2]. The goal of this study was to conduct requirements engineering for a mHealth app.

2. Methods

To achieve this, we convened 6 expert panelist groups of Spanish-speaking stakeholders from the US (N=5) and the Dominican Republic (N=6) to review the Spanish translation

[1] Corresponding Author: Rebecca Schnall RN, MPH, PhD, 560 168th Street, New York, NY 10032.

and cultural adaptation of the WiseApp. Stakeholders included clinicians, public health experts, clinic staff, and informaticians. Expert panels were conducted with a focus group format and lasted between 1-2 hours. Panelists reviewed the WiseApp and commented on the strengths and limitations of the app. Focus group questions were focused on the following areas: design, content, usability, functionality, privacy and security of the app. In addition, panelists discussed ways to simplify the language and features to accommodate Spanish-speaking users who may have lower levels of education, or color or vision deficiencies.

3. Results

Five major themes (see Table 1) emerged from these expert panels that were summarized. Findings were shared with the app developers who incorporated the feedback into an updated Spanish version of the WiseApp.

Table 1. Summary of Findings from Expert Panels

Themes	Description	Example(s)
Language/ Grammar	Syntax and sentence structure in Spanish Accurate translations in Spanish	- "manera fácil" instead of "rápida y eficiente" - "a tiempo" instead of "horario apropiado" - "administrar" instead of "manejar."
User Interface Design	Images, colors, design Formatting	- Increase button(s) size to support users with vision deficiencies - Label measurement units on tables
Usability	Features and tools to improve usability	- Provide encouraging statements to motivate users to continue taking their medication
Privacy and Security	Limitation of personal data	- Limitations of identifiers with information shared via the group forum - Login and password functions
Support	Help functions	- Videos to teach users how to use the app - A FAQ section answering common questions about the app and CleverCap unit

4. Conclusions

Findings from these stakeholder groups were integrated with findings from cognitive interviews and usability testing with end-users to improve the WiseApp and will be tested in a randomized clinical trial with Spanish-speaking patients.

References

[1] Nachega JB, Adetokunboh O, et al. Community-based interventions to improve and sustain antiretroviral therapy adherence, retention in HIV care and clinical outcomes in low-and middle-income countries for achieving the UNAIDS 90-90-90 targets. Current HIV/AIDS Reports. 2016 Oct;13:241-55.

[2] Perez-Escamilla R. Health care access among Latinos: implications for social and health care reforms. Journal of Hispanic Higher Education. 2010 Jan;9(1):43-60.

[3] Schnall R, et al. Efficacy of an mHealth self-management intervention for persons living with HIV: the WiseApp randomized clinical trial. Journal of the American Medical Informatics Association. 2023 Mar;30(3):418-26.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230191

Publication Dynamics on Social Media During the Orpea Nursing Homes Scandal: A Twitter Analysis

Mathilde FRUCHART[a,1], Laetitia VERDIER[b], Jean-Baptiste BEUSCART[a,b] and Antoine LAMER[a,c]
[a] *Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de santé et des Pratiques médicales, F-59000 Lille, France*
[b] *CHU Lille, Department of Geriatrics, F-59000 Lille, France*
[c] *F2RSM Psy - Fédération régionale de recherche en psychiatrie et santé mentale Hauts-de-France, F-59350, Saint-André-Lez-Lille, France*

Abstract. The release of a book denouncing mistreatment in French nursing home triggered a scandal which was conveyed on social networks. The objectives of this study were to study the temporal trends and dynamics of publication on Twitter during the scandal as well as to identify the main topics of discussion.The first one is spontaneous and completely aligned with the actuality and fed by media and family of residents, while the second one is out of step with current events and fed by the company involved in the scandal.

Keywords. social media, nursing home, Twitter, topic modeling, orpea, data reuse

1. Introduction

The scandal began on January 24th 2022, with the publication of a book-investigation written by the French journalist Victor Castanet. The book, untitled “Les Fossoyeurs”, revealed mistreats on elderly housed in nursing homes [1]. This scandal has provoked a political movement which denounced other nursing home groups (Korian). Social media are web based tools enable the sharing of contents. Information found in the social media may help to set up infoveillance and tracks epidemics [2, 3]. The objectives of this study were to study the temporal trends and dynamics of publication during the Orpea scandal as well as to identify the main topics of discussion on Twitter.

2. Methods

In a first step, we observed tweets related to Orpea’s scandal and we defined a lexical field referring to the scandal on Twitter as followed: "#orpeagate", “ehpad and cash investigation” (ehpad is a French acronym for nursing home), “korian”, “victor castanet”, “lesFossoyeurs”, “laurent garcia”. In a second step, we used the Tweepy library to extract

[1] Corresponding Author: Mathilde Fruchart, ULR 2694, Lille University, 2 place de Verdun, F-59000, Lille, France; E-mail: mathilde.fruchart@univ-lille.fr.

tweets including the lexical field previously defined. The extraction of tweets was done in July 2022 and covers a period of one year (from June 1st, 2021 to June 1st,2022). We applied natural language processing methods to remove duplicated tweets as well as the URLs and stopwords found in the text. We described the frequency of tweets with a timeline. We performed a topic modeling based on the unsupervised probabilistic Latent Dirichlet Allocation model (LDA). We used the Python libraries Gensim to perform topic modeling and pyLDAvis to visualize the clusters and the keywords for each topics [4]. Finally, we have extracted the posts from the Twitter account @Orpea_. The activity of the account was described by the number of posts published.

3. Results

We extracted 85,342 unique tweets related to the scandal. The main spikes concerns 10,076 tweets for the publication of "Les Fossoyeurs", 12,466 tweets for the investigation carried out by the French authorities (IGAS-IGF) and 4,494 tweets for the Cash Investigation program. Most tweets were posted between 6am and 12pm (n = 22,662, 26%).We identified 6 clusters of tweets. We manually assigned a label to each cluster. We identified 389 tweets published by the @Orpea_ account between September 2019 and November 2022. The strongest activity of the Orpea account occurred in June 2022 with 15 tweets on 01-06-2022. Conversely, there is a lack of activity on the dates of the scandal.

4. Discussion and Conclusion

The first pattern of publication on the social media is very spontaneous and aligned with actuality, with events such as (in chronological order) the release of the book "Les Fossoyeurs", the beginning of the investigation by a French authority, the senate audition of the author of the book, the TV program dedicated to another nursing home company and the complaints submitted by the state and the families of the residents. Conversely, the second pattern involved the company. It had little or no activity during the period of the scandal and did not reply to any posts. Six topics of discussion emerged: the French nursing home system, politics in general, but also denunciations of mistreatment in nursing homes, the feelings of residents' relatives, the scandal itself and the repercussions of the Orpea scandal on the other nursing home company Korian.

References

[1] *"Les fossoyeurs", Victor Castanet*. 2022. Accessed: Dec. 12, 2022. [Online].
[2] Grajales III FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. Journal of medical Internet research. 2014 Feb 11;16(2):e2912.
[3] Sun L, Tao C, Xie Z, Li D. Promotion of disposable electronic cigarette flavors and topics on twitter. International Journal of Environmental Research and Public Health. 2020 Dec;17(24):9221.
[4] Celard P, Vieira AS, Iglesias EL, Borrajo L. LDA filter: A Latent Dirichlet Allocation preprocess method for Weka. Plos one. 2020 Nov 9;15(11):e0241701.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230192

Digital Skills Among Elderly Care Workers

Eija KIVEKÄS[a,1], Ulla-Mari KINNUNEN[a], Johanna IKONEN[a] and Kaija SARANTO[a]
[a] *University of Eastern Finland, Kuopio, Finland*
ORCiD ID: Eija Kivekäs https://orcid.org/0000-0003-2440-9946, Ulla-Mari Kinnunen https://orcid.org/0000-0002-3610-3998, Johanna Ikonen https://orcid.org/0000-0002-7781-6142, Kaija Saranto https://orcid.org/0000-0002-3195-1955

Abstract. The aim of our study was to determine the current status of digital skills of elderly care workers (n=169) at well-being services. A survey was sent to elderly services providers in the municipalities (n=15) of North Savo, Finland. Respondents' experience as client information systems users was higher than that as assistive technologies users. Devices supporting independent living were seldom used, but safety devices and alarm monitoring were used daily.

Keywords. Elderly care, digital skills, client information system, assistive technology

1. Introduction

Due to demographic changes, developed countries have growing concerns about the future challenges ageing populations will present to their welfare systems [1]. Digitalization and assistive technologies (AT) have been introduced as important means of meeting these challenges. AT is used to support not only the elderly but also elderly care workers, such as: registered nurses (RN) and practical nurses (PN) [2,3]. Client information systems (CISs) are technological systems for processing, storing, and maintaining social welfare client information and documents. However, support for CISs in social work and its knowledge generation has been shown to be poor [4]. The aim of this study was to identify the digital skills of elderly care workers at well-being services in Finland.

2. Methods

An electronic questionnaire was sent to elderly care services of the municipalities (n =15/19) of North Savo autumn 2022. The questionnaire included background and multichoice questions, and three open questions. A five-point (5) Likert scale (strongly agree-strongly disagree) was used to examine variations in the use and competence of workers' digital skills. The data (n=169) was analyzed using IBM SPSS 27 to explore the respondents' CIS and AT experience, and digital skills including motivation to use digital services at work. For the analysis the original five-point scale was changed to a three-point (3) scale (agree-neither agree nor disagree-disagree).

[1] Corresponding Author: Eija Kivekäs, Department of Health and Social Management, University of Eastern Finland, E-mail: eija.kivekas@uef.fi.

3. Results

Respondents' mean age was 41 years and their working experience was wide (less than one year – more than 25 years). Nearly half (46%, n=77) worked in home care (HC), about 40% (n=70) in elderly care homes (ECH), and the rest in customer and service counselling (10%, n=16). Over half (59%, n=99) of the respondents were licensed PNs and slightly over one fifth (n=35) were RNs. Nearly 40% assessed their experience as a CIS user to be moderate (3/5, n=66) and over half assessed themselves as relatively experienced or experienced (4-5/5, n=88). Respondents were less experienced AT users (4-5/5, n=57). AT user experience was for HC workers (3.2/5) and for ECH workers (2.7/5). Respondents' digital skills and motivation can be seen in table 1.

Table 1. Digital skills and motivation*

	ECH (n=70) n / %	HC (n=77) n / %
I can search information from my organization's intranet	63 / 90	64 / 83
I know how to act as per data protection and security principles	**67 / 96**	**73 / 95**
I am motivated to develop my digital skills	61 / 87	65 / 84
I can evaluate the reliability of information I have searched	67 / 96	70 / 91
I can search for information from databases	64 / 91	67 / 87
I am motivated to use digital devices at my work	**55 / 79**	**63 / 82**
I can act according to ethical principles when I use digital services	68 / 97	71 / 92

ECH = Elderly Care Home, HC = Home Care, * = including "agree"-responses

4. Discussion

Digitalization and an increasing number of technologies aiming to support citizens' ability to live at home has raised concerns over care workers' digital skills and attitudes towards it [1,2]. Elderly care workers experience as CIS users was high, reflecting the wide implementation of electronic systems in Finnish social and health care [3,4]. In HC they used different kinds of digital devices, and they were confident with AT. A positive sign was the respondents' motivation to use digital devices and to adopt new ones.

5. Conclusion

Motivation to use digital services was high which refers to favorable situation for training digital technologies. Respondents working in HC were more confident with their CIS and AT experience than respondents in ECH. This study was funded by the European Social Fund (ESF), project code: S22852.

References

[1] Baudin K, Gustafsson C, Frennert S. Views of Swedish Elder Care Personnel on Ongoing Digital Transformation: Cross-Sectional Study. J Med Internet Res 2020;22(6):e15450. doi: 10.2196/15450.

[2] WHO 2018. Assistive technology. Available: https://www.who.int/news-room/fact-sheets/detail/assistive-technology Accessed January 2, 2023.

[3] Kivekäs E, Mikkonen S, Koponen S, Saranto K. Technology Supporting Nursing at Homecare - Seems to Be Lacking. Stud Health Technol Inform. 2020 Nov;23(275):97-101, doi: 10.3233/SHTI200702.

[4] Salovaara S, Ylönen K. Client Information Systems' Support for Case-based Social Work: Experiences of Finnish Social Workers. Nordic Social Work Research. 2022;12(3):364-78, doi: 10.1080/2156857X.2021.1999847.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230193

The Complexity of a Clinical History

Antonio D'UFFIZI[a,1], Fabrizio PECORARO[a], Fabrizio L. RICCI[a], Giuseppe STECCA[a] and Fabrizio CONSORTI[b]
[a] *Institute for Research on Population and Social Policies, National Research Council, Italy*
[b] *Italian Society of Medical Education, Italy*

Abstract. The paper describes a new metrics for measuring the structural complexity of clinical history (modelled by a HINe model) in order to compare different clinical histories and then assign it to the right types of learners.

Keywords. Complexity, Petri Nets, Health Issue, Health Issue Network

1. Introduction

Today physicians manage patients with multiple chronic diseases and need to predict their evolutions. The Health Issue Network (HIN) model describes the patient's medical history and highlights: (i) how health issues have changed over time; (ii) how affect each other [1]. This paper defines a complexity metrics for HIN graphs for the educational aim of evaluating the difficulty of an exercise based on graphs. The complexity is related to the drawing, analysis and interpretation of a HIN graph. At present, our metric does not consider the clinical complexity of the modeled case.

2. Methodology and Results

The HIN model allows to implement clinical exercises for undergraduate medical students to develop their clinical reasoning ability in multimorbidity chronic conditions [2]. The HIN model is based on the Petri Net (PN); the PN is a direct graph with two types of nodes, places (Health Issues, HI, such as diagnosis, symptom, any other clinical information), and transitions (the evolutions between HIs), linked by directed edges [3].

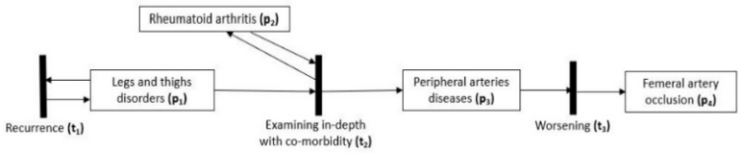


Figure 1. Example of the HINe model

Figure 1 shows a patient who develops, over time, occlusion of the femoral artery. Each HI is identified by a label p_i, while each transition is identified by a label t_i. We were inspired by the Extended Cardoso Metric [4] because it focuses on the transitions of a PN (the evolutions). In the HIN model the different types of evolutions have different

[1] Corresponding Author: D'Uffizi Antonio, E-mail: antonio.duffizi@gmail.com.

semantics and different weight in the complexity calculation. We assigned a weight also to the AND splits/joins because they influence the structural complexity of a HIN model (Table 1). These weights were inductively inferred from the errors of medical students in solving a sample of clinical exercises. The weights were finally agreed after a discussion with medical experts and are expressed in arbitrary units.

Table 1. Complexity weights

HINe Element	Weight
Worsening, Improvement, Persistence	1
Examining in-depth, Complication, Cause, AND splits/joins	2
Recurrence, Co-morbidity, Co-presence	3

Note: for the definitions of evolutions, see [2]

The complexity measuring of a HIN model is reported in the Eq. (1):

$$HIC = \sum_{p \in P} \sum_{t \in p\bullet} W_t(t) + \sum_{a \in A} W_a(a) + |P| \quad (1)$$

- P and T, the finite set of places/HIs and transitions/evolutions, with $P \cap T = \emptyset$;
- A, the finite set of AND splits/joins, with $A \subseteq T$;
- W: $T \rightarrow \mathbb{N}$, the function of the complexity weight (see Table 1);

The formula $\sum_{t \in p\bullet} W_t(t)$ calculates the total weight of transitions/evolutions, output of place p ($t \in p\bullet$). For the HIN model depicted in Figure 1, the Eq. (1) becomes:

$$\boldsymbol{HIC} = \left[W_t(t_1)|_{t_1 \in p_1\bullet} + W_t(t_2)|_{t_2 \in p_1\bullet}\right] + W_t(t_2)|_{t_2 \in p_2\bullet} + W_t(t_3)|_{t_3 \in p_3\bullet} + W_a(a) * |\{t_2\}| + |\{p_1, p_2, p_3, p_4\}| = [3+2] + 3 + 1 + (2*1) + 4 = \mathbf{15}$$

Next, this HIC can be compared to the HICs of other clinical histories to establish the clinical history with the most complex structure.

3. Discussion and Conclusions

The assessment of the difficulty of an exercise allows to students to be provided with exercises with a progressive cognitive load. Future works will consider the clinical difficulty for an overall evaluation of complexity of a clinical case. A wider evaluation by medical students and doctors (professors, doctoral students, residents, physicians) in agreement with the Italian Society for Medical Education (SIPeM) is ongoing.

Acknowledgments: Authors wish to thank the Italian Society of Medical Education (SIPeM), professors and students from "Sapienza" University of Rome, and Daniela Luzi and Oscar Tamburis for the useful discussion.

References

[1] Ricci FL, et al. A Petri-Net-Based Approach for Enhancing Clinical Reasoning in Medical Education. IEEE Transactions on Learning Technologies. 2022;15(2):167-78. doi: 10.1109/TLT.2022.3157391.
[2] Health Issue Network. Available from: https://www.healthissuenetwork.org/eng/home.
[3] Reisig W. A Primer in Petri Net Design: Springer: Berlin/Heidelberg, Germany; 2012. 132 p.
[4] Lassen KB, van der Aalst WMP. Complexity metrics for Workflow nets. Information and Software Technology. 2009;51(3):610-26. doi: 10.1016/j.infsof.2008.08.005

This page intentionally left blank

Section 4

Decision Support Systems

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230195

The Effect of Physicians' Acknowledgement of Clinical Decision Support Systems Generated Alerts on Patient Diabetes Management in a Primary Care Setting

Mohammad A. FAYSEL[a,1], Tameca MILLER[b], Jonathan SINGER[c] and Caroline CUMMINGS[c]
[a]*Health Informatics Program, School of Health Professions, SUNY Downstate Health Sciences University, Brooklyn, NY, USA*
[b] *The Brooklyn Hospital Center, Brooklyn, NY, USA*
[c]*Department of Psychological Sciences, Texas Tech University, Lubbock, Texas, USA*
ORCiD ID: Mohammad A. Faysel https://orcid.org/0000-0002-5584-9074

Abstract. The purpose of the study is to evaluate whether clinician's acknowledgement and adherence to Clinical Best Practice Advisories (BPA) system's alerts improves the outcome of patients with chronic diabetes. We used deidentified clinical data of elderly (65 or older) diabetes patients with hemoglobin A1C (HbA1C) >= 6.5 that were extracted from the clinical database of a multi-specialty outpatient clinic that also provides primary care services. We performed paired ttest to evaluate whether clinician's acknowledgement and adherence to BPA system's alert has any impact on patients' HbA1C management. Our findings showed that the average HbA1C values improved for patients whose alerts were acknowledged by their clinicians. For the group of patients whose BPA alerts were ignored by their clinicians, we found clinicians' acknowledgement and adherence to BPA alerts for chronic diabetes patient management did not have a significant negative effect on improvement in patient outcome.

Keywords. Clinical Decision Support System, CDSS and Diabetes Management, Acknowledgement of CDSS alerts

1. Introduction

The use of Clinical Decision Support System(CDSS) tools in primary care settings has increased dramatically in recent years, especially for the use of managing patients with chronic conditions [1]. CDSS are computer based information systems that help physicians, nurses and other healthcare professionals with medical decision making. CDSS has been found to have a profound impact on improving physician behavior and doctor performance [1]. A team-based care for diabetes management using CDSS for Medicaid patients showed a significant reduction in patients' HbA1C [2]. A systematic

[1] Corresponding Author: Mohammad Faysel, E-mail: Mohammad.faysel@downstate.edu.

review and meta-analysis on the effect of CDSS on cardiovascular risk factors revealed that complying with CDSS generated guidelines resulted in significant clinical benefits among patients with type 2 diabetes in attaining target low density lipoprotein cholesterol (LDL-c) [3]. A comparative study examining the effectiveness of team-based care with and without a CDSS for diabetes management showed a significant reduction in the cardiovascular risk factors in diabetes patients when the team-based care employed and used a CDSS [4]. Another study explored the use of CDSS for diabetes management in Scotland and found CDSS to have positive impact on improving patients' HbA1C [5]. A systematic review by Sly et al. found CDSS to have a significant positive impact on safety and quality of inpatient diabetes management [6]. However, numerous studies investigating the efficacy of CDSS at improving patient outcome, safety and cost of care are lacking positive findings [1, 7-8].

Studies that target specific diseases (e.g., diabetes; hypertension) have been sparse and the limited results have been equivocal in nature. There have been a few studies that have found CDSS to be effective in managing and improving the process of care for type 2 diabetes mellitus [9-10]. In a recent systematic review that investigated the effects of CDSS in individuals with type 2 diabetes mellitus, Clevering and colleagues concluded that CDSS was effective in improving the process of care, adding feedback on performance and/or case management [10]. However, this systematic review found equivocal results in regard to it effect on decreasing glucose levels. It should be noted that Tsai, Wang, Hsu & Li [11] conducted a study to investigate false positive alerts and whether they affect outcomes when investigating CDSS's effectiveness. They concluded that most studies do not differentiate the true positive and false positive alerts, and therefore, the effects of true positives and false positive alerts will be mixed. The implication of this is an underestimation of how effective CDSS is in reducing the negative effects of type 2 diabetes. This also increases the likelihood of Type II error when investigating CDSS.

The purpose of the study is to evaluate whether clinician's acknowledgement and adherence to Clinical Best Practice Advisories (BPA) system's alerts, a type of CDSS, embedded within an electronic health record (EHR) system improves the outcome of patients with chronic diabetes. The alert logic was designed in a way that it fired an alert based on a patient's condition to help the provider to acknowledge and review the patients chronic care plan for that year. The BPA's criteria were built for patients that had a chronic condition of diabetes or chronic heart failure and had not been seen for evaluation during the encounters. We hypothesized that the greater use of the BPA system by practitioners, the better the outcome for the patient suffering from chronic diabetes, as demonstrated by lower HbA1C levels.

2. Methods

For this retrospective study, deidentified clinical data for one month of elderly (65 or older) diabetes patients with hemoglobin A1C (HbA1C) >= 6.5 were extracted from the clinical database of a multi-specialty outpatient clinic that also provides primary care services. Average age of patients was 79 with a standard deviation 8.9. The dataset consists of 64% female and 36% male patients.

The dataset was collected from a multidiscipline outpatient practice. The practice served the population within the boroughs of New York City excluding the Bronx. It included a team of 100 primary care physicians and an estimated 4 support staff per

provider that comprised of a clerk, a medical assistant, a nurse and a mid-level provider such as a physician assistant or a nurse practitioner. The primary care physician along with support staff support a model called a provider care team. This model aligned in line with Patient Centered Medical Home (PCMH). PCMH is a model that has been created to allow organizations to structure their healthcare delivery system to best suit the healthcare providers and the patients. This model was geared to allow primary care facilities to provide efficient and cost effective care without compromising the quality and coordination of care a patient received.

For each of these patients, a CDS alert based on BPA indicating elevated HbA1C was generated that required primary care clinician's attention and further clinical intervention(s). To comply with the widely used National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) guidelines, the organization's protocol requires repeating HbA1C test every 3 months for diabetes patients and every 6 months for patients whose blood sugar is well-controlled. We used patient records whose alerts were acknowledged, and the records with ignored alerts to verify whether the clinician's acknowledgement to BPAs improves chronic diabetes patient outcome.

For the group whose alerts were acknowledged, we claim ed that due to clinician's acknowledgement of the alerts, average HbA1C value after alerts would be improved (lowered) compared to the average of the last HbA1C value s prior to the alerts. Our null hypothesis for this case was that the average HbA1C would remain unchanged or worsen (elevated) after clinicians' acknowledgement of the alerts. A paired ttest was performed to verify our claim. For the records whose alerts were ignored, we also claimed that the average HbA1C value after the alerts would be worsen (elevated) compared to the average of their last HbA1C values prior to the alerts due to clinicians' ignoring of the alerts. Our null hypothesis for this case was that the average HbA1C would remain unchanged or improved (lowered) before and after the alerts . We performed a paired ttest to verify our claim for this group. All the data analyses were performed, and statistical values and 95% Confidence Intervals (CI) were generated using SAS 9.4 software. This study protocol was approved by SUNY Downstate Health Sciences University's IRB.

3. Results

Out of total 264 alerts, 191 (72.3%) were acknowledged by clinicians and 73 (27.7%) were ignored. For the group whose alerts were acknowledged, the paired ttest produced a t value of 2.9 with a probability value of 0. 00 4 (p-value < 0.05) which is sufficient to reject the null hypothesis and supported our claim. It also generated a n HbA1C mean difference of 0.24 (95% CI, 0.08 - 0.41) between before and after acknowledgement of the alerts.

For the group whose alerts were ignored, the paired ttest produced a t value of 1.87 with a probability value of 0.0 7 (p - value > 0.05) which supported the null hypothesis for this group and rejected our claim that the average HbA1C value of this group after the alert would be worsen (elevated) compared to the average of their last HbA1C values prior to the alerts. This test also showed that the average HbA1C improved (lowered by 0.17) after alerts compared to the average of the last HbA1C values prior to alert (95% CI, - 0.0 1 - 0.36).

4. Discussion

As the rates of diabetes mellitus continue to increase, the financial burden of the disease is also increasing as there has been a 41% increase in the United States from 2007 to 2012 alone; costing $174 billion to $245 billion [12]. Therefore, decreasing the rate of diabetes mellitus through the use of inexpensive interventions is warranted and can help to mitigate or even prevent other chronic illnesses. One such possible intervention is CDSS, as it has been demonstrated to be cost-effective and can be used to serve a large number of individuals with a generally low cost. Technology vendors, healthcare organizations and researchers have been working collaboratively to identify effective use of CDSS for chronic disease management. Our study finding has shown that clinicians' acknowledgement of CDSS alerts (BPA alerts) for diabetes patient management has a positive impact on patients' diabetes improvement (lower average HbA1C value) which is consistent with findings from recent literature review [9].

On the other hand, our study has also revealed that the group of patients whose BPA alerts were ignored by their clinicians did not have their average HbA1C value increase after the alerts were ignored. This could be due to improper labeling of BPA alerts in the system or false alerts or alert fatigues that caused the clinicians to ignore the alerts. This could also be due to clinicians ignoring the alerts because they were false alarms and just treating the patients based on standard protocol set forth by the organization without acknowledging them in the system.

5. Limitations

Our study had some limitations and some of the findings sound inconsistent with findings from current literature. For example, our study showed clinicians' ignoring of BPA alerts did not have any significant negative effect on the improvement in patient outcome which contradicted some of the findings from current literature [13]. The dataset lacked information on patient demographics such as race and ethnicity, clinician's expertise and training such as whether a clinician were properly trained on how to use the CDSS and take appropriate actions based on the system generated BPA alerts. Due to sensitivity of practice related business protocol, organizational requirement on BPA alerts and how these CDSS tools governed the clinical practice protocol to improve patient outcome were not available in the dataset.

The authors plan to explore comprehensive data that include patient level information, detailed information on clinicians, and clinical practice settings to address these limitations of this study and to identify factors associated with clinicians' ignoring CDSS generated alerts that could have clinical impact on the management of chronic conditions in primary care settings. However, knowing the reasons behind clinicians' lack of willingness to adhere to the CDSS generated alerts or ignoring the alerts in the primary care clinical settings can have profound impact on designing and optimizing CDSS tools that could potentially meet end user clinicians' needs and improve adherence to clinical protocols for chronic diabetes management in technology-enabled clinical settings.

6. Conclusion

The study has shown that the average HbA1C values improved for patients whose alerts were acknowledged by their clinicians. This finding aligns with findings from most of the studies that focused on CDSS and diabetes management [13]. The group of patients whose BPA alerts were ignored by their clinicians, our analysis showed that clinicians' acknowledgement and adherence to BPA alerts for chronic diabetes patient management did not have any significant negative effect on the improvement in patient outcome. This could be due to the study dataset's lack of detail information on clinicians' adherence to organizational protocols. The authors would like to explore this in the future to address this issue using larger datasets from multiple clinical sources that include both patient and physician level information.

References

[1] Murphy ME, Fahey T, Smith SM. Computer-based clinical decision support for general practitioners. Fam Pract. 2014 Oct;31(5):497-8. doi: 10.1093/fampra/cmu045. Epub 2014 Aug 5. PMID: 25096276.

[2] Zhang X, Svec M, Tracy R, Ozanich G. Clinical decision support systems with team-based care on type 2 diabetes improvement for Medicaid patients: A quality improvement project. Int J Med Inform. 2021 Nov 18;158:104626. doi: 10.1016/j.ijmedinf.2021.104626. Epub ahead of print. PMID: 34826757.

[3] Groenhof TKJ, Asselbergs FW, Groenwold RHH, Grobbee DE, Visseren FLJ, Bots ML; UCC-SMART study group. The effect of computerized decision support systems on cardiovascular risk factors: a systematic review and meta-analysis. BMC Med Inform Decis Mak. 2019 Jun 10;19(1):108. doi: 10.1186/s12911-019-0824-x. PMID: 31182084; PMCID: PMC6558725.

[4] Shi X, He J, Lin M, Liu C, Yan B, Song H, Wang C, Xiao F, Huang P, Wang L, Li Z, Huang Y, Zhang M, Chen CS, Obst K, Shi L, Li W, Yang S, Yao G, Li X. Comparative Effectiveness of Team-Based Care With and Without a Clinical Decision Support System for Diabetes Management: A Cluster Randomized Trial. Ann Intern Med. 2023 Jan;176(1):49-58. doi: 10.7326/M22-1950. Epub 2022 Dec 6. PMID: 36469915.

[5] Conway N, Adamson KA, Cunningham SG, Emslie Smith A, Nyberg P, Smith BH, Wales A, Wake DJ. Decision Support for Diabetes in Scotland: Implementation and Evaluation of a Clinical Decision Support System. J Diabetes Sci Technol. 2018 Mar;12(2):381-388. doi: 10.1177/1932296817729489. Epub 2017 Sep 14. PMID: 28905658; PMCID: PMC5851216.

[6] Sly B, Russell AW, Sullivan C. Digital interventions to improve safety and quality of inpatient diabetes management: A systematic review. Int J Med Inform. 2022 Jan;157:104596. doi: 10.1016/j.ijmedinf.2021.104596. Epub 2021 Oct 2. PMID: 34785487.

[7] Heselmans, A., Delvaux, N., Laenen, A. et al. Computerized clinical decision support system for diabetes in primary care does not improve quality of care: a cluster-randomized controlled trial. Implementation Sci 15, 5 (2020). https://doi.org/10.1186/s13012-019-0955-6

[8] Ilag LL, Martin CL, Tabaei BP, Isaman DJ, Burke R, Greene DA, Herman WH. Improving diabetes processes of care in managed care. Diabetes Care. 2003 Oct;26(10):2722-7. doi: 10.2337/diacare.26.10.2722. PMID: 14514570.

[9] Jia P, Zhao P, Chen J, Zhang M. Evaluation of clinical decision support systems for diabetes care: An overview of current evidence. J Eval Clin Pract. 2019 Feb;25(1):66-77. doi: 10.1111/jep.12968. Epub 2018 Jun 26. PMID: 29947136.

[10] Cleveringa FG, Gorter KJ, van den Donk M, van Gijsel J, Rutten GE. Computerized decision support systems in primary care for type 2 diabetes patients only improve patients' outcomes when combined with feedback on performance and case management: a systematic review. Diabetes Technol Ther. 2013 Feb;15(2):180-92. doi: 10.1089/dia.2012.0201. PMID: 23360424.

[11] Tsai CY, Wang SH, Hsu MH, Li YC. Do false positive alerts in naïve clinical decision support system lead to false adoption by physicians? A randomized controlled trial. Comput Methods Programs Biomed. 2016 Aug;132:83-91. doi: 10.1016/j.cmpb.2016.04.011. Epub 2016 Apr 13. PMID: 27282230.

[12] American Diabetes Association. (2013). Economic costs of diabetes in the US in 2012. Diabetes Care, 36(4), 1033-1046.

[13] Ali SM, Giordano R, Lakhani S, Walker DM. A review of randomized controlled trials of medical record powered clinical decision support system to improve quality of diabetes care. Int J Med Inform. 2016 Mar;87:91-100. doi: 10.1016/j.ijmedinf.2015.12.017. Epub 2015 Dec 30. PMID: 26806716.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230196

Secur-e-Health Project: Towards Federated Learning for Smart Pediatric Care

Rita RB-SILVA [a,1], Xavier RIBEIRO [b,c], Francisca ALMEIDA [b,c], Carolina AMEIJEIRAS-RODRIGUEZ [a], Julio SOUZA [a,d], Luis CONCEIÇÃO [b], Tiago TAVEIRA-GOMES [a,c], Goreti MARREIROS [b] and Alberto FREITAS [a,d]
[a] *Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Porto, Portugal*
[b] *GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, LASI – Intelligent Systems Associate Laboratory, Institute of Engineering – Polytechnic of Porto, Porto, Portugal*
[c] *MTG Research and Development Lab, Porto, Portugal*
[d] *CINTESIS@RISE, Faculty of Medicine, University of Porto, Porto, Portugal*
ORCiD ID: Rita Rb-Silva https://orcid.org/0000-0002-1422-0974; Xavier Ribeiro https://orcid.org/0000-0002-5316-0732; Francisca Almeida https://orcid.org/0000-0002-3872-0682; Carolina Ameijeiras Rodríguez https://orcid.org/0000-0001-8817-7772; Julio Souza https://orcid.org/0000-0002-8576-1903; Luis Conceição https://orcid.org/0000-0003-3454-4615; Tiago Taveira-Gomes https://orcid.org/0000-0002-0998-6000; Goreti Marreiros https://orcid.org/0000-0003-4417-8401; Alberto Freitas https://orcid.org/0000-0003-2113-9653.

Abstract. The application of machine learning (ML) algorithms to electronic health records (EHR) data allows the achievement of data-driven insights on various clinical problems and the development of clinical decision support (CDS) systems to improve patient care. However, data governance and privacy barriers hinder the use of data from multiple sources, especially in the medical field due to the sensitivity of data. Federated learning (FL) is an attractive data privacy-preserving solution in this context by enabling the training of ML models with data from multiple sources without any data sharing, using distributed remotely hosted datasets. The Secur-e-Health project aims at developing a solution in terms of CDS tools encompassing FL predictive models and recommendation systems. This tool may be especially useful in Pediatrics due to the increasing demands on Pediatric services, and the current scarcity of ML applications in this field compared to adult care. Herein we provide a description of the technical solution proposed in this project for three specific pediatric clinical problems: childhood obesity management, pilonidal cyst post-surgical care and retinography imaging analysis.

Keywords. federated learning; machine learning; privacy-preserving protocols; clinical decision support; pediatrics.

[1] Corresponding Author: Rita Rb-Silva, E-mail: rrsilva@med.up.pt

1. Introduction

In recent years, Federated Learning (FL) has gained increased attention from the academic, medical and data science communities due to its capability of training artificial Intelligence (AI) models using data from multiple sources while maintaining data privacy [1]. This Machine Learning (ML) paradigm enables collaborative algorithm training without exchanging or retrieving data from its sources, thereby addressing many of the existing barriers for achieving the full potential of AI applications in healthcare, namely those related to data governance and privacy [2]. Moreover, recent literature indicates that FL models can achieve performance levels that are comparable to those trained on centrally hosted datasets, and even superior to those estimated from isolated single-institutional datasets [3].

Amongst the plethora of clinical applications of AI models obtained through FL, there is still lack of research focusing on pediatric care, despite the current workforce challenges and rising demands on pediatric services due to increased patient complexity and comorbidity [4]. In this context, the Secur-e-Health project aims at developing a solution in terms of Clinical Decision Support (CDS) tools encompassing FL predictive models and recommendation systems for smart pediatric care. In this paper, we describe the technical solution proposed within the Secur-e-Health project for three specific pediatric clinical problems (childhood obesity treatment, pilonidal cyst post-surgical management and retinography imaging analysis).

2. Methods

A full CDS solution has been designed according to functional and technical requirements defined with the help of clinicians in order to satisfy current healthcare needs and gaps considering three specific clinical scenarios. Clinicians' feedback were obtained from virtual periodic meetings with two physicians and a nurse working in a Portuguese teaching hospital.

In the first scenario, related to overweight and obesity management in children, it is intended to build a predictive medical decision support system to aid clinicians in developing nutritional and exercise recommendations to maximize patient adherence in the long term and to anticipate circumstances that may reduce treatment effectiveness, considering the aspects of the child and its direct family socioeconomic, education and health conditions. In the second scenario, related to pilonidal cyst post-surgical care, a predictive medical decision support system is intended to predict optimal frequency visits following surgery and to estimate the optimum resource allocation for post-surgical treatments. In the third and last scenario, regarding smart retinography images analysis, a set of predictive models is proposed to detect relevant clinical phenotypes or comorbidities by means of retinal fundus imaging analysis, that would otherwise be identified later in the natural history of child development or non-communicable diseases.

3. Results

3.1. Proposed architecture

Secur-e-Health system, being a federated learning model, is composed by two main elements: an aggregator and the remote distributed training sites (Figure 1). These two elements communicate through a study package, *i.e.* an algorithm that runs through the distributed data in a private and secure fashion, to determine a weight gradient (W) that will be used to feed the equation:

$$W_{t+1} = W_t + f(W^k)$$

and give the global results.

The main users of Secur-e-Health will be healthcare professionals. Specifically, the system will train predictive ML models with relevant data from EHR and provide the healthcare professionals with feedback and recommendations to maximize the adherence of overweight/obesity patients to the therapeutic plan, as well as to optimize the post-surgical care of patients with pilonidal cysts. Additionally, the healthcare professionals will be also provided with the output from predictive ML models that analyze retinography images and integrate the extracted quantitative features with clinical information.

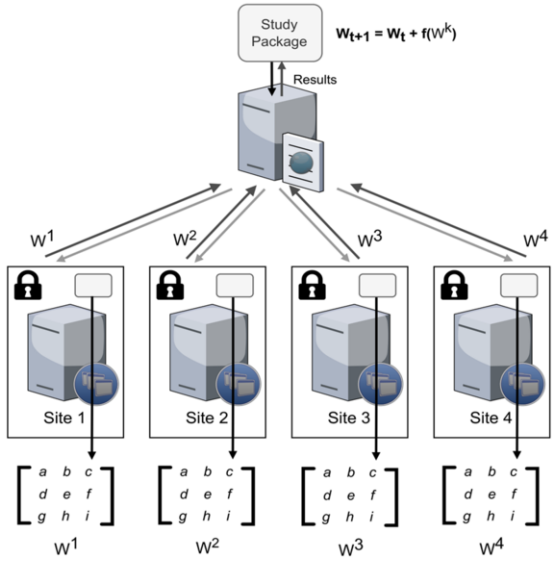


Figure 1. General architecture of the Secur-e-Health system.

3.2. Requirements

The business and technical requirements shared by all the use cases are displayed on Table 1. Table 2 displays the specific business and technical requirements for each use case.

Table 1. Transversal business and technical requirements for the three pediatric use cases.

Transversal business requirements	Transversal technical requirements
1. To analyze structured and unstructured data from EHRs; 2. To assess and improve the quality of data through a process of data cleansing; 3. To display personalized feedback for healthcare professionals; 4. To ensure the privacy and security of patient data.	1. Data gathering: the algorithm should be able to gather information from EHRs; 2. Data quality: the algorithm should be able to ensure data quality, by fixing inaccurate, incomplete, inconsistent and repetitive data, by removing duplicates and by adding the data necessary to complete the records; 3. Decision support: the algorithm should be able to incorporate all kind of characteristics, process them and output results in a natural medical language that help professionals to make decisions regarding a specific condition; 4. Privacy and security compliance: data used to train the models should be used complying with privacy and security norms concerning patient and family data; 5. Data access control: server software application programming interface (API) should provide user authentication and authorization.

Table 2. Business and technical requirements for each use case addressing a specific pediatric clinical problem.

Use case	Business requirements	Technical requirements
Childhood overweight and obesity management	1. To identify and gather clinical information regarding pediatric overweight/obesity from EHRs; 2. To identify and gather information on lifestyle choices, family socioeconomic status, education level, eating habits and general health conditions; 4. To integrate clinical information with individual lifestyle choices and family context; 5. To use ML models to profile patient and family features; 6. To develop a CDS application to aid clinicians in developing recommendations; 7. To promote patient and/or caregivers' adherence to the lifestyle therapy plan; 8. To implement quality control over time and to use ML to improve the effectiveness of the CDS application.	1. Data integration: the algorithm should find associations between clinical features and individual lifestyle choices and family context; 2. Individual feedback: the algorithm should output individual results per patient, in order to provide feedback to the named client and maximize the adherence of patients and/or caregivers to the therapy plan; 3. Quality control: the algorithm should constantly control its own quality and improve the effectiveness of the CDS application.
Pilonidal cyst surgery (PCS)	1. To identify and gather relevant clinical information from EHRs regarding PCS; 2. To identify and gather information on education level, eating habits and general health conditions; 3. To identify and gather relevant clinical information on the surgical technique; 4.To identify and gather clinical information regarding follow-up visits; 5. To integrate clinical information with individual lifestyle choices, hereditary factors, issues related to the chosen surgical technique and anesthesia, and events at follow-up visits; 6. To develop predictive models (ML algorithms) to support medical decisions related to PCS; 7. To develop a stand-alone CDS system;	1. Data integration: the algorithm must be able to integrate different types of variable values, regarding patients, at any point in time; 2. Data selection: the algorithm will filter the data that are relevant relative to patients submitted to SC surgery; 3. Model improvement: the algorithm must continually access the prediction results to constantly improve some of the parameters to be used; 4. Clinical assistance: the algorithm must output individual

	8. To promote patient and/or caregivers' adherence to post-surgical care of PCS.	results in order to maximize patient adherence and optimize the scheduling of further appointments.
Retinal imaging analysis	1. To process retinography images; 2. To automatically assess the main factors interfering with the analysis of retinography images and to compare that assessment with the human assessment; 4. To identify the main anatomical features of the retina and to assess the blood vessels segmentation using retinography images; 5. To determine, gather and provide clinical quantitative features from retinography images; 6. To extract and gather morphometric features, comorbidities and other relevant clinical information from the EHRs; 7. To integrate quantitative features of retinographies with clinical, laboratory and/or medical imaging data; 8. To create predictive models for the presence of relevant clinical phenotypes or comorbidities; 9. To compute metrics for the treatment plan and goals assessment; 10. To use ML models to profile patient and family features and to help with the medical decision.	1. Image analysis: the algorithm must analyze the retinography images and extract quantitative features, as well as assess factors interfering with the analysis; 2. Assessment similarity: the algorithm must be able to compare its automatic assessment with the human visual assessment; 3. Model improvement: supervised ML should be used to extract transferrable features from retinography data to increase model performance; 4. Predictive models: the algorithm must use techniques like linear regression, multiple regression, logistic regression, decision trees, random forests, data mining, and neural networks.

4. Conclusions

The main goal of the proposed solution in Secur-e-Health is to optimize adherence to medical recommendations in pediatric care, and thus achieve greater treatment success rates, and healthcare cost savings in the long term. The proposed solution of Secur-e-Health will provide: (i) secure and privacy preserving cross-organizational analysis by providing MultiParty Computation, Federated Learning techniques and AI-driven algorithms; (ii) use of predictive models integrating a medical decision support system to aid pediatricians, namely in the scenarios of childhood obesity management, sacrococcygeal cyst surgical management and integration of complex retinal fundus imaging data to enrich the aforementioned models; (iii) further validation of the algorithms into AI-enabled digital care pathways.

References

[1] Dayan I, Roth HR, Zhong A, *et al*. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021 Oct; 27(10):1735-1743. doi: 10.1038/s41591-021-01506-3

[2] Xu J, Glicksberg BJ, Su C, *et al*. Federated Learning for Healthcare Informatics. J Healthc Inform Res. 2021 Mar; 5(1): 1-19. doi: 10.1007/s41666-020-00082-4

[3] Rieke N, Hancox J, Li W, *et al*. The future of digital health with federated learning. NPJ Digit Med. 2020 Sep 14;3:119. doi: 10.1038/s41746-020-00323-1

[4] Clarke SL, Parmesar K, Saleem MA, Ramanan AV. Future of machine learning in paediatrics. Arch Dis Child. 2022 Mar;107(3):223-228. doi: 10.1136/archdischild-2020-321023

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230197

PDSS: A Pharmacological Decision Support System for Diabetics Patients with COVID-19

Isabel AMAYA-RODRIGUEZ[a,1], Nekane LARBURU[b], María Rollán MARTINEZ-HERRERA[c], Kristin REBESCHER[b], Iván MACIA[b], Miguel A. ARMENGOL DE LA HOZ[d], Cristina RUBIO-ESCUDERO[a] and Alba GARIN-MUGA[b]

[a] *Department of Computer Science, University of Seville, Spain*

[b] *Biodonostia Health Research Institute, (Bioengineering Area), eHealth Group, San Sebastián 20014, Spain*

[c] *Applied Chest Imaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

[d] *Big Data Department - Regional Ministry of Health of Southern Spain, Spain*

ORCiD ID: Isabel Amaya-Rodríguez https://orcid.org/0000-0002-1073-7116, Nekane Larburu https://orcid.org/0000-0003-0248-7783, María Rollán Martínez-Herrera https://orcid.org/0000-0002-4667-521X, Kristin Rebescher https://orcid.org/0000-0002-8169-9655, Iván Macia https://orcid.org/0000-0003-0448-7840, Miguel A. Armengol de la Hoz https://orcid.org/0000-0002-7012-2973, Cristina Rubio-Escudero https://orcid.org/0000-0001-5732-9139, Alba Garín-Muga https://orcid.org/0000-0002-7160-1191

Abstract. With the advent of SARS-CoV-2, several studies have shown that there is a higher mortality rate in patients with diabetes and, in some cases, it is one of the side effects of overcoming the disease. However, there is no clinical decision support tool or specific treatment protocols for these patients. To tackle this issue, in this paper we present a Pharmacological Decision Support System (PDSS) providing intelligent decision support for COVID-19 diabetic patient treatment selection, based on an analysis of risk factors with data from electronic medical records using Cox regression. The goal of the system is to create real world evidence including the ability to continuously learn to improve clinical practice and outcomes of diabetic patients with COVID-19.

Keywords. Diabetes; Clinical decision support systems; COVID-19; Risk Factor

1. Introduction

On January 30th, 2020, the World Health Organization (WHO) declared a public health emergency of international concern due to coronavirus disease (COVID-19). COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and is characterised by its strong transmission capacity, and although most patients experience mild symptoms or moderate illness, according to the WHO, 10-15% of cases progress to

[1] Corresponding Author: Isabel AMAYA-RODRIGUEZ, E-mail: iamaya@vicomtech.org.

severe disease. As of December 15th, 2022, there were 646,740,524 confirmed COVID-19 cases [1] and approximately 26% of detected COVID-19 patients have Diabetes Mellitus (DM) [2] depending on the study. Due to the high incidence of diabetics with COVID-19, several studies have analysed and reported on the negative impact that DM may have on patients admitted to the hospital with COVID-19 based on survival analysis (SA) [2]. However, to the best of our knowledge, no article in the literature proposes the development of a pharmacological decision support system for DM patients with COVID-19 based on SA. In this article, we propose a recommendation system based on risk factors attending to their impact on survival of various drugs administered DM patients with COVID-19 treated in HM Hospitales (HM), a private hospital group based in Spain. The results obtained may be useful in clarifying the optimal pharmacological protocol for COVID-19 diabetes patients. Proposed analysis methodologies may also apply to other PDSS in other disease areas.

2. Methods

Given the unprecedented nature of the pandemic and the wealth of information held within electronic medical records, HM, recognized early-on the importance of making anonymised COVID-19 data publicly available. Through its "Covid Data Saves lives" project, HM published anonymised clinical, demographic, and laboratory data for the first time on April 15th 2020 to identify trends and narrow in on establishing a standard of care for COVID-19 patients more rapidly. The study was approved by the Ethics committee of the HM Hospitals (approval number 22.04.2005-GHM). After access granting, the raw data (N=2561) was cleaned to improve its completeness and usability removing patients with missing values or outliers and those cases without confirmed COVID-19 diagnosis or incomplete data, which were approximately 21% of the total. The final cohort included 2,101 patients diagnosed with SARS-Cov-2 infection. The variables in the dataset included demographic information, laboratory analyses, comorbidities, and administered pharmaceutical treatment as part of their care during their hospital stay, among others.

The purpose of this study is to evaluate the effectiveness and factors affecting outcomes for drugs administered to COVID-19 patients with DM. Therefore, patients that had the following ICD-10 diagnosis codes (E11.9, E11.5, E11.22, E11.319, and O24.410) were considered patients with DM, which were 18,18% of the total. Additionally, we selected the patients who were administered the most common pharmaceutical treatments against COVID-19 such as Hydroxychloroquine, Azithromycin, Azithromycin/Hydroxychloroquine combination, Lopinavir/Ritonavir combination, Antiviral HIV, Tocilizumab, Oseltamivir, Interleukin 1 inhibitors, Interferons, Glucocorticoids, Dexamethasone, Hydrocortisone, Heparine and Organic Nitrate.

A preliminary extraction of features was performed to simplify the variables that were collected from the EHR of the patients involved in the analysis. The variables selected in this pre-processing were chosen following clinical criteria and those that had statistical significance ($p<0.005$) between diabetic and non-diabetic patients in the data set. The characteristics analysed were age, sex, hospitalization days (1-7 days, 8-14 days, 15-30 days, 30-90 days, 90+ days), discharge status (home, voluntary, other hospital, social-sanitary centre, death), comorbidities and ICU admission (yes/no).

Once these variables were identified, the analysis then focused on obtaining which drugs improve or worsen the survival of diabetics with COVID-19 and what are the influencing factors for survival. Thus, given these influencing factors, potential hypothesis to generate real world evidence (RWE) in the form of clinical rules were obtained following the procedure of evidence-based clinical guideline development [3]. Next, we present the components of our PDSS for DM patients with COVID-19.

2.1. Pharmacological Decision Support System Components

The PDSS consists of two main components: i) the Risk Factors Discovery Engine that analyses the key risk factors related to the different treatments for COVID-19 in diabetic patients and ii) the Pharmacological Rule System, that making usage of the key factors formalizes new knowledge and, based on input clinical information from the patient's EHR, returns as a result a set of drug recommendations or contraindications for a diabetic patient [4].

2.2. Risk Factors Discovery Engine

As mentioned in the methods, the factors included in the risk analysis were obtained from the *preliminary extraction* to guide the identification of the relevant factors that affect the drug administration. They were selected by applying the different statistical tests mentioned below using a value of $p < 0.05$ as the threshold applied to determine statistical significance. Categorical data were analysed using the Chi-square test and in the case of numerical data, Fisher's exact test or the Mann-Whitney U test was applied.

Once the selected variables were included, we need to identify which are the influencing factors to generate the hypothesis for our study. The following steps were carried out:

1. For each drug:
    a. Diabetic patients were divided into two groups, those who had received the drug and those who had not.
    b. Survival analysis with Kaplan-Meier curves was performed with both groups and the Log Rank test was applied to demonstrate differences between cohorts when $p<0.05$ (Table 1).
2. If differences were demonstrated (1.b.), Cox regression analysis was performed to look for factors influencing patient survival (Table 1)

When the hypothesis is validated with enough evidence (RWE), the corresponding rules are implemented in the Pharmacological Rule System (PRB), which incorporate the relevant risk factors obtained (2).

2.3. Pharmacological Rule System

Once the risk factors were identified, RWE in the form of digital rules needed to be implemented, following similar approaches to guideline-based CDSS systems [4]. First the subjects that are involved in the study need to be identified. In this case DM patients with COVID-19. Next, the PRS is able to receive as input the relevant patient clinical EHR data, collect the data from the variables obtained in the *preliminary extraction* and

return whether the drug may be recommended for the patient or not. Additionally, the system may evolve and automatically create new RWE based rules improving the knowledge base [4]. The following image shows the workflow of the PDSS system:

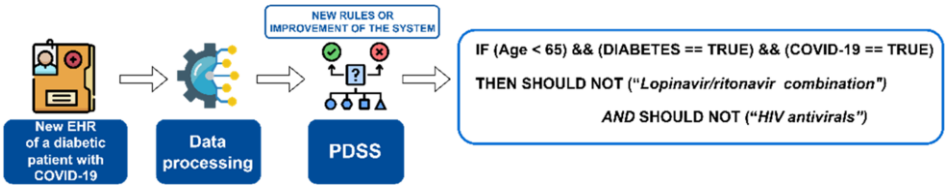


Figure 1. PDSS workflow

3. Results

Table 1 includes the significative results of the survival analysis comparison in patients with DM who received or not a treatment after applying Kaplan-Meier and Log-Rank. As result, the drugs with relevant impact in the survival rate of diabetics were Hydroxychloroquine, Azithromycin, Azithromycin/Hydroxychloroquine combination, Lopinavir/Ritonavir combination, Antiviral HIV and Heparin.

The resulting *preliminary extraction* variables were the age, the hospitalization days between 1-7 days and several comorbidities such as Hypertension, Obesity, Allergies, Hyperlipidemia/Dyslipidemia, Apnoea, Chronic Kidney Disease, Chronic Heart Disease and Acute Renal Failure. Despite of sex and ICU admission not being statistically significative in the *preliminary extraction*, they were added in the risk factors discovery engine due to previous studies showed that they are important facts in the survival of COVID-19 patients and due to this it was added [5,6]. The age was divided in two groups <65 years and >= 65 years as was applied by [2].

Table 1. Survival and risk factors results by treatment

Drug Name	N(%) Diabetics with the treatment	Log Rank[1] (p value)	Factors do not Improve Survival[2]	Factors improving Survival[3]
Hydroxychloroquine	308 (83.6)	<0.005	-	ICU and Hypertension and Allergies
Azithromycin	215 (58.4)	.04	-	ICU and Hypertension and Allergies
Azithromycin / Hydroxychloroquine combination	202 (54.8)	<.005	-	ICU and Hypertension and Allergies
Lopinavir/ritonavir combination	154 (41.8)	.03	Age < 65 years	ICU and Hyperlipidemia /Dyslipidemia and Acute Renal Failure
Antiviral HIV	157 (42.6)	.02	Age < 65 years	ICU and Hyperlipidemia /Dyslipidemia and Acute Renal Failure
Heparine	10 (84.2)	<0.005	Sex == Female	ICU and Allergies

[1] Log Rank (p value) for DM patients who had received the drug and those who had not; [2] Factors do not Improve Survival obtained in COX regression analysis; [3] Factors improving Survival obtained in COX regression analysis

The results presented in Table 1 could be transformed to multiple hypothetic rules form which the conditions would be the identified risk factors. As an example, below one potential hypothetic RWE-based rule for a patient under 65 years with COVID-19 and DM is presented:

IF (Age < 65) && (DIABETES == TRUE) && (COVID-19 == TRUE)
THEN SHOULD NOT ("*Lopinavir/ritonavir combination")*
AND SHOULD NOT ("*HIV antivirals")*

4. Discussion and Conclusion

We have proposed a methodology to obtain hypothesis for COVID-19 patients with DM based on a preliminary analysis of risk factors affecting outcomes for given treatments. This methodology allows to generate RWE on cases based on statistical analysis using survival analysis (Kaplan-Meier and Cox regression) and translate this evidence into rule-based decision support systems. The development of such a system is relevant in scenarios such a pandemic or health crisis situations, where evidence needs to be generated quickly in absence of further evidence from clinical trials. Similar approaches may be followed for other diseases in real-world data applications such as survival analysis in value-based healthcare scenarios. As future work, the recommendations obtained will be validated by clinical professionals. Besides, further variables should be analysed to improve the risk factors recommendations. For example, allergies or specific variables related to comorbidities, such as Creatinine in the case of Acute Renal Failure. Finally, data-driven approaches, such as machine learning models, could be incorporated for complex data scenarios, to complement statistical approaches in outcome prediction.

5. References

[1] WHO Coronavirus (COVID-19) Dashboard n.d. https://covid19.who.int (accessed December 17, 2022).
[2] Shang J, Wang Q, Zhang H, Wang X, Wan J, Yan Y, et al. The Relationship Between Diabetes Mellitus and COVID-19 Prognosis: A Retrospective Cohort Study in Wuhan, China. Am J Med 2021;134:e6–14. https://doi.org/10.1016/j.amjmed.2020.05.033.
[3] Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing guidelines. BMJ 1999;318:593–6.
[4] Torres J, Artola G, Muro N. A Domain-Independent Semantically Validated Authoring Tool for Formalizing Clinical Practice Guidelines. Stud Health Technol Inform 2020;270:517–21. https://doi.org/10.3233/SHTI200214.
[5] Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. Nature 2020;588:315–20. https://doi.org/10.1038/s41586-020-2700-3.
[6] Roncon L, Zuin M, Rigatelli G, Zuliani G. Diabetic patients with COVID-19 infection are at higher risk of ICU admission and poor short-term outcome. J Clin Virol 2020;127:104354. https://doi.org/10.1016/j.jcv.2020.104354.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230198

Parralel Recurrent Convolutional Neural Network for Abnormal Heart Sound Classification

Arash GHAREHBAGHI[a,1], Elaheh PARTOVI[b] and Ankica BABIC[c,d]
[a] *School of Information Technology, Halmstad University, Halmstad, Sweden*
[b] *Department of Electrical Engineering, Amirkabir University, Tehran, Iran*
[c] *Department of Biomedical Engineering, Linköping University, Linköping, Sweden*
[d] *Department Information Science and Media Studies, University of Bergen, Bergen, Norway*

Abstract. This paper presents the results of a study performed on Parallel Convolutional Neural Network (PCNN) toward detecting heart abnormalities from the heart sound signals. The PCNN preserves dynamic contents of the signal in a parallel combination of the recurrent neural network and a Convolutional Neural Network (CNN). The performance of the PCNN is evaluated and compared to the one obtained from a Serial form of the Convolutional Neural Network (SCNN) as well as two other baseline studies: a Long- and Short-Term Memory (LSTM) neural network and a Conventional CNN (CCNN). We employed a well-known public dataset of heart sound signals: the Physionet heart sound. The accuracy of the PCNN, was estimated to be 87.2% which outperforms the rest of the three methods: the SCNN, the LSTM, and the CCNN by 12%, 7%, and 0.5%, respectively. The resulting method can be easily implemented in an Internet of Things platform to be employed as a decision support system for the screening heart abnormalities.

Keywords. Heart sound, deep learning, parallel convolutional neural network, convolutional neural networks, intelligent phonocardiography.

1. Introduction

Recent progress in development of different deep learning methods, created a leap towards intelligent decision making in various domains of healthcare including medical informatics and biomedical engineering. Extraction of medical information from time series of physiological activities has been traditionally regarded as an essential domain of research, sometimes with vital importance. Heart sound signal analysis, so called intelligent phonocardiography [1–4], is one of the highlighted domains of medical informatics and biomedical engineering which has been increasingly receiving attentions from the researchers, particularly after development of the powerful deep learning methods such as Convolutional Neural Network (CNN) and Long- and Short Term Memory (LSTM) [5–6]. The number of the publications addressing this domain has been doubled during the last three years with respect to the preceding half decade, mainly due to the high potential of deep learning methods in learning subtle details of the time series.

[1] Corresponding Author: Arash Gharehbaghi, E-mail: arash.ghareh-baghi@hh.se.

Among the various deep learning methods, spanning from the ones sophisticated for heart sound analysis, i.e. Time Growing Neural Network [7–9], to the image-based learning methods, CNN is by far the most popular learning method seen in different publications on heart sound analysis [5–7]. Although serious criticisms addressed reliability of the presented methods in terms of the structural risk [3,7,10], the study performed based on cascading a CNN and a recurrent neural network, which we named SCNN [5], and is still considered as the state-of-the art of this topic. It is evident that a reliable machine learning method can be incorporated into a digital stethoscope to serve as an easy-to-use and inexpensive screening tool for detecting heart abnormalities, which is, yet, considered as a priority for primary healthcare centers.

This paper introduces a Parallel combination of CNN and LSTM for heart sound signal analysis (PCNN), and compares accuracy of the presented method to the ones obtained by Deng et al. [5] in discriminating between normal and abnormal heart sounds. Performance of two other baseline methods were explored for the comparison: a Conventional CNN (CCNN) and a LSTM. Accuracy of the 4 methods is statistically validated using the widely known public dataset of heart sound, the PhysioNet/Computing in Cardiology Challenge 2016 (https://physionet.org). Results of the study published by Deng et al. are still considered as the state-of-the-art in terms of the performance and the methodology. This was the main motivation for selecting the Computing in Cardiology Challenge as the baseline for comparison.

2. Materials and Methods

The abovementioned repository of heart sound recordings contains 6 folders of data, named training-a to training-f, altogether comprising 3240 signals, from which 665 signals correspond to abnormal hearts. The subjects can have multiple recordings; however, the recordings have been allocated in the folders in a mutually exclusive manner. The recordings are all anonymous and have different lengths from 5 to 20 seconds. More details are found in (https://physionet.org/content/challenge-2016/1.0.0/).

2.1. Classification Methods

The classification methods are trained using the datasets exist in the mentioned repository, in which the number of the normal signals are by far higher than the abnormal ones, yielding a heavy class imbalance in the training data. In the PCNN case, an input signal is firstly divided into nonoverlapping segments of 5 seconds. In order to overcome the class imbalance, heavily seen in the repository, an augmentation method of SMOTE was employed. Details of the augmentation method can be found in [5]. The signal contents are mapped to 2-dimensional representation using the mel-frequencies representation [5]. Contents of the mel-frequency are employed by a CNN and a LSTM independently and the ultimate classification is performed using the two sets of the outcomes. In another attempt, a cascaded connection of CNN and LSTM, named CCNN, was employed in which the LSTM performs the ultimate classification. The rest of the processing including the mel-frequency representation remain identical to the PCNN. Figure 1 demonstrates the block diagram of the two methods. In both of the cases, the CNN and the LSTM are independently trained and optimized using Adam optimizer. The set of the hyper parameters is identically selected for the PCNN and the CCNN as listed in Table 1. The baselines for comparison are composed of a CCNN and a LSTM

with the identical inputs of mel-frequency contents together with the similar set of the hyperparameters as the ones selected for the PCNN, Figure 1.

Table 1. The hyperparameters

Parameter	Value
Kernel size and stride size of convolution layer	3×3 and 1
Number of the convolutional layers and kernel at each layer	3 and (16, 32, 64)
Number and size of the Max Pool	3 and ((2×2), (4×4), (2×2))
Number of LSTM layers and units	1 and 74
Number of neurons in FC and the dropout rate	32 and 0.5
Activation function of the last layer	SoftMax
Activation function of the convolutional and FC layers	ReLu
Initial learning rate and exponential decay rates	0.01 and (0.9, 0.999)
Batch size and number of the epoch	512 and 50

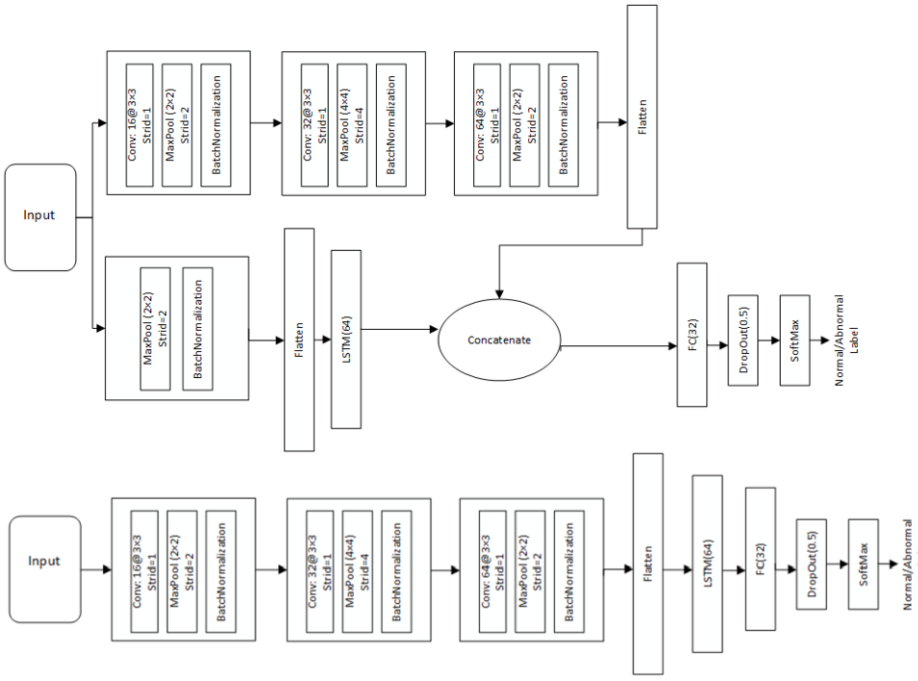


Figure 1. Block diagram of the PCNN (top) and CCNN (bottom)

3. Results

The recordings of each folder are employed for evaluation of the 4 methods, the PCNN, the SCNN, the CCNN, and the LSTM, with the training/validation/test of 75%/15%/10% selected from the Physionet datasets. The signals of each Physionet folder are firstly shuffled. Then, the training/validation/test recordings are selected according to the mentioned percentages. The three methods are independently evaluated 5 times and the descriptive statistics of two performance measures, the accuracy and the sensitivity, are calculated. Table 2 lists the estimated values for the average and standard deviation

(STD) of each performance measure, independently. As can be seen from the table, the PCNN offers the best performance in terms of the accuracy. The CCNN and LSTM both exhibit better accuracies comparing to the SCNN, however, this is not true with the sensitivities which are substantially lower with high standard deviations. This makes the CCNN and LSTM inappropriate to be employed as a screening tool. On the other hand, accuracy of the SCNN shows a high standard deviation against the evaluation data (the training/validation/test data), implying on its high structural risk. Nevertheless, for a certain selection of the evaluation data, an outperformance of SCNN might be observed, which could not be assumed objectively conclusive for the general population.

Table 2. The descriptive statistics of the performance measures for the three methods, PCNN, CCNN, and CNN for 5 runs of the algorithms with the train\validation\test of 75%\10%\15%

Method	Accuracy		Sensitivity	
	Average (%)	STD (%)	Average (%)	STD (%)
PCNN	87.2	2.7	76.5	14.9
SCNN	75.2	25.1	76.6	16.0
CCNN	86.7	2.6	64.5	21.6
LSTM	80.3	2.9	32.3	17.3

4. Discussion

The paper suggested a parallel combination of LSTM and CNN for discriminating between normal and abnormal heart sounds. This combination not only improves performance of the conventional CNN and LSTM, but also provides a better accuracy comparing to the cascaded combination which is considered as the state-of-the-art. The sensitivity is not improved by the parallel combination, though. The cascaded combination of CNN and LSTM was previously introduced by Deng et al. [5], with a very high accuracy. However, this is limited to a specific selection of the training/validation/test datasets, and thus very data-sensitive leading to a conclusion that cannot not be generalized as reflected by the high standard deviation of the performance measures. An important aspect of this study is the richness of dataset for training and testing of the methods. This richness is somewhat impaired by the class imbalance for the normal and abnormal heart sounds. In this study, an augmentation method, named SMOT was employed to overcome the class imbalance. Nonetheless, a large dataset with sufficient samples of the normal and abnormal samples, covering various pathological conditions, is crucial to arrive at an optimal selection for training the hyperparameters. The resulting method has the practical potential to improve the screening accuracy of cardiac auscultation at the primary healthcare centers. Studies showed that this screening accuracy is, yet, insufficient [11,12]. Integration of such artificial intelligence-based methods assigns a high level of sophistication to the electronic stethoscopes towards associating intelligence to the stethoscopes for various diagnostic objectives [13].

5. Conclusions

The parallel combination of LSTM and CNN improves performance of a cascaded LSTM and CNN for discrimination between normal and abnormal heart conditions using the heart sounds. This parallel combination offers significant enhancement of performance in terms of accuracy as compared to the conventional cases of CNN and LSTM, showing capability of such a combination of the deep learning methods as employed in this case study. However, the gain in accuracy could not guarantee an increase of the sensitivity or its reproducibility. The proposed machine learning method with high and stable accuracy can be integrated with an electronic stethoscope to be employed as a decision support system at primary healthcare centers.

References

[1] Gharehbaghi A, Sepehri AA, Lindén M, Babic A. Intelligent Phonocardiography for Screening Ventricular Septal Defect Using Time Growing Neural Network. Informatics Empowers Healthcare Transformation: IOS Press; 2017. 238:108-11.

[2] Gharehbaghi A, Lindén M. An Internet-Based Tool for Pediatric Cardiac Disease Diagnosis Using Intelligent Phonocardiography. In: Mandler B, Marquez-Barja J, Mitre Campista ME, Cagáňová D, Chaouchi H, Zeadally S, et al., editors. Internet of Things IoT Infrastructures: Second International Summit, IoT 360° 2015, Rome, Italy, October 27-29, 2015 Revised Selected Papers, Part I. Cham: Springer International Publishing; 2016. p. 443-7.

[3] Gharehbaghi A, Babic A. Structural Risk Evaluation of a Deep Neural Network and a Markov Model in Extracting Medical Information from Phonocardiography. Studies in health technology and informatics, 2018, 251:157-160.

[4] Gharehbaghi A, Sepehri AA, Lindén M, Babic A, A Hybrid Machine Learning Method for Detecting Cardiac Ejection Murmurs. EMBEC & NBC 2017; 787-790.

[5] Deng M, Meng T, Cao J, Wang S, Zhang J, Fan H. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Netw. 2020 Oct;130:22-32.

[6] Singh SA, Meitei TG, Majumder S. Short PCG classification based on deep learning. InDeep Learning Techniques for Biomedical and Health Informatics 2020 Jan 1 (pp. 141-164). Academic Press.

[7] Gharehbaghi A, Babic A. Deep Time Growing Neural Network vs Convolutional Neural Network for Intelligent Phonocardiography. Stud Health Technol Inform. 2022 Jun 29;295:491-494. doi: 10.3233/SHTI220772. PMID: 35773918.

[8] Gharehbaghi A, Lindén M, Babic A. An Artificial Intelligent-Based Model for Detecting Systolic Pathological Patterns of Phonocardiogram based on Time-Growing Neural Network. Applied Soft Computing, 2019. 83:105615.

[9] Gharehbaghi A, Lindén M. A Deep Machine Learning Method for Classifying Cyclic Time Series of Biological Signals Using Time-Growing Neural Network. IEEE Transactions on Neural Networks and Learning Systems. 2018; 29(9) 4102-4115.

[10] Gharehbaghi A. and Babic A., A-Test for Quantifying Structural Risk and Learning Capacity of Supervised Machine Learning Methods, IOS Press, 2022.

[11] Watrous RL, Thompson WR, Ackerman SJ. The impact of computer‐assisted auscultation on physician referrals of asymptomatic patients with heart murmurs. Clinical Cardiology: An International Indexed and Peer‐Reviewed Journal for Advances in the Treatment of Cardiovascular Disease. 2008 Feb;31(2):79-83. https://doi.org/10.1002/clc.20185

[12] DeGroff CG, Bhatikar S, Hertzberg J, Shandas R, Valdes-Cruz L, Mahajan RL. Artificial neural network–based method of screening heart murmurs in children. Circulation. 2001 Jun 5;103(22):2711-6..

[13] Gharehbaghi A, Ask P, Nylander E, Janerot-Sjoberg B, Ekman I, Lindén M, Babic A. A hybrid model for diagnosing sever aortic stenosis in asymptomatic patients using phonocardiogram. InWorld Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada 2015 (pp. 1006-1009). Springer International Publishing.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230199

Clinical Decision Support: Evaluating the Development of a Tool for Nurses

Bente CHRISTENSEN[a,b,1] and Ann Kristin ROTEGÅRD[c]
[a] *The Norwegian Nurses Organisation, Norway*
[b] *Nord University, Norway*
[c] *VAR Healthcare, Cappelen Damm, Norway*

Abstract. VAR Healthcare is a clinical decision support system for nurses that aspires to become even more advanced. By applying The Five Rights model, we have evaluated the status and direction of its development to bring potential lacks or barriers into the fore. The evaluation shows that ensuring APIs that will allow the nurses to combine the assets of VAR Healthcare with information on individual patients from EPRs would bring advanced decision support to nurses. This would adhere to all the principles of the five rights model.

Keywords. Clinical Decision Support, knowledge support, Five Rights model, electronic patient record systems, documentation of nursing, evidence-based procedures, evidence-based guidelines.

1. Introduction

Health care has become extremely complex. There are more complex patient situations, numerous professions involved, complex information flow, and higher turnover of patients. Additionally, treatment and nursing has become extremely advanced, exposing health care personnel to loads of knowledge and an increasing scope of new research that should be made available to and implemented into practice. Thus, there is a need for, and an expectation that digital tools would assist health care personnel in keeping an overview of their patients and providing updated treatment- and care regimes.

The benefits of Electronic Patient Records (EPR) have been increasingly recognized in healthcare. If Clinical Decision Support (CDS) is provided within EPRs based on practice guidelines, care can be provided based on evidence and best practice to achieve better quality and efficiency [1-5]. Yet another promising potential for CDS is the ability to close the gap between clinical practice, research, and education [6]. However, there is a lack of a general CDS that can support nurses across patient conditions, and of systems that are fully integrated with the EPR systems [5,7]. Research on CDS utilization has mostly targeted physicians, with little mention of nurses [8]. The studies of CDS that exists in nursing are mostly tools within the acute care settings and for single clinical conditions, e.g., pressure ulcers haemodynamic instability, respiratory distress, and infection detection [9]. Also, there is a lack of studies that includes CDS development, and a lack of usability studies [7].

[1] Corresponding Author: Bente Christensen, E-mail: bente.christensen@nsf.no

The core of decision support for nurses is evidence-based interventions with connected nursing diagnosis and patient outcomes. Specifically, nurses need support to define and document nursing diagnosis and these should be predefined and anchored in research and provided automatically [8].

Besides standardized, structured data, decision support presupposes a knowledge base that can be combined with the data on individual patients [10,11].

Despite high expectations of quality achievements, it has been difficult to bring decision support systems into routine use in clinical practice. Four characteristics have been identified that significantly contributed to improve clinical practice; the CDS is automatically available in the clinical workflow, the support is provided through the system at the time and place (site) of the decision, provides practical advice / references, and are computer-based [4]. CDS that is an integrated component of the patient record or its order entry system, are significantly more likely to improve clinical practice than standalone systems [3,12].

VAR Healthcare (VAR) started as a public research and development project and is a well-established and still growing CDS system for nurses and is being widely used across the continuum of care and patient conditions in Norway and Denmark the last 20 years. Because the usage is covering primary and secondary care, as well as education in nursing and healthcare this tool closes the gap between clinical practice, research and education. This system consists of evidence-based procedures (interventions) with belonging knowledge summaries (rationales), nursing diagnosis and goals. The content is continuously updated and is integrated to the EPR systems. Furthermore, it is based on various standards and structures, e.g. the ISO, International Organization for Standardization, 9000), and a conceptual model based on four key concepts: well-being, integrity, prevention, and safety (VIPS). VIPS is a structure and keywords for representation of nursing care in patient records. [13,14].

VAR also utilizes reference terminology (International Classification for Nursing Practice - ICNP) to ensure interoperability, an advanced search engine and the possibility to further develop sustainable integrations and information flow across systems [15,16]. A few research projects have involved VAR. However, VAR seems to respond to the gaps in research and development. Therefore, our aim was to evaluate VAR as a decision support tool by using the Five rights model [12] as a framework.

2. Methods

We used a case study approach with a formative evaluation of the VAR tool. The aim was to evaluate and document the development and use of VAR Healthcare and ensure that the tool was evolving in the right and meaningful direction for an advanced clinical decision support tool for nurses.

We used the CDS Five Rights model as a framework. This framework has been widely used for such analysis and is considered a best practice approach for quality improvement and healthcare outcomes, when the interventions in the CDS transfer the right information, to the right person, in the right format, through the right channel at the right time in the workflow [12].

The right information means evidence-based information that provide guidance or advice for relevant, best practice. Examples may be clinical guidelines, procedures or pathway templates.

The right person refers to the person who need the support for best possible process and outcome. This can be health care professionals (physicians, nurses, pharmacist, etc), patients, and/or their next of kin/caretakers.

The right format refers to how the support is given i.e. data displays, documentation tools, care plans, registry reports.

The right channels might be a clinical information system like the electronic health record system, or more general channels like Internet, mobile technology systems, smart home devices and patient portals.

The right time refers to the timing of the support or guidance in the clinical workflow or –process.

The analysis used 1. the described steps in VAR Healthcare's development and its components, 2. the gaps in CDSSs described in literature [6-8], 3. in depth information of the tool in terms of a synthesis of findings from a previous case study covering a) literature review of studies and publications with VAR, b) e-mail inquiries to the VAR team c) interviews with super users of the VAR tool [17]. The findings of these three main elements were mapped to the Five Rights principles/components.

Two researchers conducted the evaluation independently and met for comparison and discussions until consensus was reached.

3. Results

The right information. The core in VAR Healthcare are procedures for nurses that are anchored in research and best evidence/practice, and furthermore that is used across settings in healthcare. The knowledge base in VAR is rooted in science, personal and collective experience, as well as professional traditions.

Also, systematic and structured use of nursing theories and models are part of the foundation. This is line with the findings in literature of CDS in nursing, where authors suggest use of theories to improve the realizations of CDS in nursing practice [6]. VAR has been used across health care settings and education through 20 years, hence the practical knowledge and experience-based knowledge has developed cumulatively based on input from experts across settings, regions, and countries. The previous case study uncovered the users wishes for more procedures to be developed and this is put into system in a continuous process of development and maintenance in VAR.

The right person. VAR Healthcare is developed for nurses, by nurses. Increasing and strengthen nurses' competences is an important assumption to strengthen patient safety, continuity of care and positive patient outcomes. Not only providing the right content to the right person but providing the pedagogical perspective and research findings have the aim of supporting nurses to provide and understand the rational for their actions. This aims at increased reflection over practice and professionality.

The right format. VAR provides evidence and knowledge in a practical manner and language, and illustrations are in concordance with the text also through the up-dates. Technology has been utilised to provide various levels of information based on the user's need, e.g., a nursing student or a nurse who needs more in-depth information can open a full vision of the main steps of the procedure, while experienced nurses might only need the main steps of the procedure or just a glance of a detailed animation or illustration. A grading system of changes and knowledge tests that guides the nurse to find the right answer, ensures lifelong learning and help nurses keep up to date as an integrative part of their work. Technologically the tool has a responsive design and adjusts to the various

devices, has been found easy to use and user-friendly. Information from the case study report uncovered that the knowledge summaries was less used.

The right channels. VAR is web-based and has a responsive design, meaning that VAR can be accessed from different devices, like the computer, smartphones and tablets, on Android and IOs devices. The access is over IP, so the nurses does not have to log in, but they can register as personal users for access from their private devises too. Perhaps most important is that it can be reached through integrations in EPRs. Literature shows that CDS as an integrated component of the patient record or its order entry system, were significantly likely to improve clinical practice than standalone systems [3,12].

The right time. Integrations with EPR systems also enables nurses to have decision support in their clinical workflow, at the time they need it. The advanced search engine allows for quick and easy access to actual procedures in a busy daily life, included support to find (remember) what interventions to perform in various patient situations (problem and risk diagnosis). The use of ICNP in VAR will be of importance for building further support, including serving nurses information at the right time, as certain information about the individual patient might trigger certain procedures, nursing diagnoses or goals (expected outcomes) and can be served to the nurses automatically. This will enable nurses to act on real-time information, instead of acting retrospectively.

4. Concluding Discussion

The evaluation uncovered that VAR qualifies well based on the Five Rights. A general knowledge system, using standard terminology that ensures interoperability, to be used across continuum of care and across countries/languages, has not been described in the literature. Also, the system is in continuous development. The findings from the case study used in the evaluation shows that the nurses are satisfied with the tool and the services provided. However, one part of the content was found to be less used: the knowledge summaries declaring the rationale for the procedures. This information is important to strengthen in clinical practice to make nurses realise and become aware of why they are promoting actions, be confident in providing best practice to patients, and improve collaboration within and across health professions. It should be considered whether a re-ux design is needed to support the nurses better in using this part of the tool.

Also, the Five rights model can be used to strengthen the implementation and use of VAR in practice. To ensure continuous, clinical care of good quality, ensure patient safety and best possible health outcomes, data must be structured and must be of good quality. Clinical Data models (CDM) and reference terminologies/-models are needed to ensure even semantic interoperability. The advanced search engine in VAR serves as a decision support tool and is evolving in the right direction to become an advanced clinical decision support system for nurses across settings and countries. However, attention should be given to "the right time" of the rights in the Five Rights model. To utilize the benefits from structured data, clinical data models and the reference terminology, close collaboration with EPR vendors should be established. There is a need to develop deeper integration, e.g. based on rest-call APIs with the EPR systems, to provide nurses with automatic suggestions based on a combination of information on individual patients from the EPR and evidence-based suggestions from VAR. Such integrations would bring CDS to the next level that could further optimize clinical practice and documentation – the next level of decision support for nurses. Nurses will be able to practice nursing care

of the highest quality, based on evidence, and more efficiently, which in turn contributes to integrated care pathways and patient safety, as well as better use of the resources.

Moreover, the combination of high-quality knowledge base and real time patient data from the EPR, may ensure better quality data for statistics and research.

References

[1] Hyruk LA. Information Systems and Decision Support Systems: What are they and how are they used in nursing? AJN. 2012 112(1):62-65

[2] Wang P, Zhang H, Baohua L, Lin K. Making Patient Risk Visible: Implementation of a Nursing Document Information System to Improve Patient Safety. In Sermeus W. et al. (Editors). IMIA and IOS Press. 2016 Doi: 10.3233/978-1-61499-658-3-8.

[3] Barretto S, Warren J, Goodchild A, Bird L, Heard S, Stumptner M. Linking guidelines to Electronic Health Record design for improved chronic disease management. AMIA Symposium. 2003 66-70.

[4] Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. MBJ. 2005 330: 765

[5] Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann, H, Tang PC. Clinical Decision Support Systems for the Practice of Evidence-based Medicine. J. Am. Med. Inform. Assoc. 2001 8: 527–534.

[6] Van der Lei J. Closing the loop between clinical practice, research, and education: the potential of electronic patient records. Methods Inf. Med. 2002 41: 51–54

[7] Lopez K, Febretti A, Stifter J, Johnson A, Wilkie D, Keenan G. Toward a More Robust and Efficient Usability Testing Method of Clinical Decision Support for Nurses Derived From Nursing Electronic Health Record Data. International Journal of Nursing Knowledge. 2017 28(4): 211-218

[8] Borum C. Barriers to hospital-based nurse practitioners utilizing clinical decision support systems. CIN Computers Informatics Nursing. 2018 36(4):1

[9] Bennett P, Hardiker NR. The use of computerized clinical decision support systems in emergency care: a substantive review of the literature. J Am Med Inform Assoc. 2017 May 1;24(3):655-668. doi: 10.1093/jamia/ocw151

[10] Beslutningsstøtte: Definisjoner, status og forvaltning av ulike former for IKT-basert klinisk støtte. Rapport fra ekspertgruppe. Versjon 1.0. Oslo: Helsedirektoratet; 2014: https://www.nsf.no/Content/1527347/Beslutningsst%C3%B8tte%20rapport%202014_v1%200%20%282%29.pdf Accessed June 15, 2022

[11] Goossen-Baremans A, Collins S, Park HA. Semanticification in connected health. Stud Health Technol Inform. 2017 232:133-151.

[12] Osheroff JA, Teich JA, Levick D. et al. Improving Outcomes with Clinical Decision Support: An Implementer's Guide. 2nd Edition. Chicago, IL: HIMSS, 2012. American Medical Informatics Association. 2017 24(3): 669-676.

[13] Ehrenberg A, Ehnfors M, Thorell-Ekstrand I. Nursing documentation in patient records: experience of the use of the VIPS model. Journal of Advanced Nursing. 1996 24 (4): 853-867.

[14] Ehnfors M, Ehrenberg A, Thorell-Ekstrand I. Nya VIPS-boken. Välbefinnande, integritet, prevention, säkerhet. 2. utg. Lund: Studentlitteratur; 2013

[15] ISO/TS 13972; 2015 (en). Health informatics – Detailed clinical models, characteristics and processes https://www.iso.org/obp/ui/#iso:std:iso:ts:13972:ed-1:v1:en, Accessed 24.07.22

[16] WHO family of classifications: https://unstats.un.org/unsd/classifications/Family/Detail/1024 Accessed June, 15, 2022

[17] Rotegård AK, Fossum M. Fra oppslagsverk til beslutningsstøtte–VAR Healthcare som case. Digitalisering i sykepleietjenesten–en arbeidshverdag i endring. 2019 p.177-200

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230200

Prediction of COVID-19 Mortality in the Intensive Care Unit Using Machine Learning

Aikaterini SAKAGIANNI[a], Christina KOUFOPOULOU[b], Vassilios VERYKIOS[c], Evangelos LOUPELIS[d], Dimitrios KALLES[c] and Georgios FERETZAKIS[c,e,1]
[a] *Sismanogleio General Hospital, Intensive Care Unit, Marousi, Greece*
[b] *Aretaieio Hospital, National and Kapodistrian University of Athens, Anesthesiology Department, Athens, Greece*
[c] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[d] *Sismanogleio General Hospital, IT department, Marousi, Greece*
[e] *Sismanogleio General Hospital, Department of Quality Control, Research and Continuing Education, Marousi, Greece*

Abstract. Since its emergence, the COVID-19 pandemic still poses a major global health threat. In this setting, a number of useful machine learning applications have been explored to assist clinical decision-making, predict the severity of disease and admission to the intensive care unit, and also to estimate future demand for hospital beds, equipment, and staff. The present study examined demographic data, hematological and biochemical markers routinely measured in Covid-19 patients admitted to the intensive care unit (ICU) of a public tertiary hospital, in relation to the ICU outcome, during the second and third Covid-19 waves, from October 2020 until February 2022. In this dataset, we applied eight well-known classifiers of the caret package for machine learning of the R programming language, to evaluate their performance in forecasting ICU mortality. The best performance regarding area under the receiver operating characteristic curve (AUC-ROC) was observed with Random Forest (0.82), while k-nearest neighbors (k-NN) were the lowest performing machine learning algorithm (AUC-ROC: 0.59). However, in terms of sensitivity, XGB outperformed the other classifiers (max Sens: 0.7). The six most important predictors of mortality in the Random Forest model were serum urea, age, hemoglobin, C-reactive protein, platelets, and lymphocyte count.

Keywords. Artificial intelligence; machine learning; COVID-19; SARS-CoV-2; ICU—intensive care unit.

[1] Corresponding Author: Georgios FERETZAKIS, Ph.D., School of Science and Technology, Hellenic Open University, Patra 263 35, Greece, E-mail: georgios.feretzakis@ac.eap.gr.

1. Introduction and Background

The 2019 coronavirus disease (COVID-19) pandemic has posed an ongoing threat to global health since its emergence. As of 9 December 2022, 643,875,406 confirmed cases of COVID-19, including 6,630,082 deaths, have been reported to WHO [1]. One of the biggest challenges facing healthcare systems around the world during the Covid-19 pandemic is the surge in patient numbers and concomitant limited medical resources. Several recent studies suggest that machine learning-based predictive models of disease severity and outcome could meet these requirements in the best possible way permitting efficient allocation of resources in the intensive care unit [2-7]. However, no sufficiently validated prognostic models are currently in wide clinical use.

In this study, we aim to compare the performance of eight different machine learning methods for predicting the mortality of critically ill COVID-19 patients admitted to the Intensive Care Unit (ICU), using a simple dataset containing demographic information and results of routine laboratory tests. We also aim to identify the most important variables associated with ICU mortality using machine learning methods.

2. Methods and Materials

A retrospective observational study was performed in the Intensive Care Unit (ICU) of a tertiary public hospital during the second and third waves of the COVID-19 pandemic. The study was approved by the Institutional Review Board (IRB) of Sismanogleio General Hospital (no. 4333/2022). For this analysis, we used the caret package [8] in R [9], a well-known framework for building ML models. The caret package provides functions to support model training for complex classification and regression tasks. During a 17-month period from October 2020 until February 2022, all the patients with PCR-confirmed COVID-19 pneumonia that were admitted to the ICU were included in the study. The primary outcome was ICU mortality, and the secondary outcome was the identification of predictive variables for mortality.

We investigated patient demographic data, blood biomarkers routinely measured in Covid-19 patients admitted to the ICU, bloodstream infection occurrence, and ICU outcome. Analytically, the data set includes the following variables: age, gender, Procalcitonin (PCT), Creatine Kinase (CPK), Troponin, Creatinine (CREA), C-Reactive Protein (CRP), Lactate Dehydrogenase (LDH), serum levels of Urea (UREA), Ferritin (Ferr), serum glutamate-pyruvate transaminase (SGOT), serum glutamate-oxaloacetate transaminase (SGPT), total bilirubin (TBIL), activated partial thromboplastin time (aPTT), D-Dimer, International Normalized Ratio (INR), hemoglobin (HGB), white blood cells count (WBC), lymphocyte count (LYM%), neutrophil count (NEUT%), platelets (PLT), the presence of bloodstream infection (microbial or/and fungal, none) and the binomial ICU outcome (survival or death).

2.1. Data preprocessing

We used the laboratory values of the two first days of the ICU stay. Then after taking the first 7 observations in chronological order from each patient, we take the average in each variable and with this value, we replace the missing values and after that, we keep only the first two lines from each patient.

2.2. Model Building

The classification problem that we had to deal, with consisted of two classes of patients; the patients who survived and were discharged alive from the ICU and the patients who died during their ICU stay. We applied and evaluated the following classification models: linear discriminant analysis (method: lda), Recursive Partitioning and Regression Trees (method: rpart), k–nearest neighbor (method: knn), support vector machines (method: svmRadial), random forests (method: rf), eXtreme Gradient Boosting (method="xgbLinear"), AdaBoost Classification Trees (method="ada"), and Stochastic Gradient Boosting (method="gbm") to try to predict hospitalization. The caret package uses the train() function to build any predictive model. We trained our models through a 10-fold cross-validation (CV) procedure to avoid overfitting in our analysis.

2.3. Model Evaluation

The performance metrics of the ML techniques that were evaluated in this study were the Sensitivity, Specificity, and Area Under the Receiver Operating Characteristic Curve AUC ROC from prediction scores.

3. Results

Among 373 patients (140 female, 233 male) with COVID-19 pneumonia that were included in the study, 102 (27.34%) died in the ICU. Patients who died were older (mean [SD] age 70 yr [11.6 yr] vs 64.4 yr [13.5 yr]) and more likely to be female (28.6% vs 26.6%). The descriptive statistics for Age, Gender, and Outcome are summarized in Table 1. The best performance considering the area under the receiver operating characteristic curve (AUC-ROC) was observed with Random Forest (0.82), followed by XGB (0.80), while k-nearest neighbors (k-NN) were the lowest performing machine learning algorithm (0.59). In terms of sensitivity, XGB outperformed the other classifiers, with maximum sensitivity reaching 0.7. The corresponding AUC-ROC curves of the eight classifiers in the specific data set are shown in Figure 1. The relative importance of the examined variables with Random Forest sorted in descending order is shown in Figure 2. The six highest-ranked variables were Urea, Age, Hgb, CRP, PLT, and lymphocyte count.

Table 1. Descriptive statistics for Age, Gender, and Outcome

Age (years)		Gender (%)			Outcome (%)
Median	68	Male	233 (62.46)	Survived	371 (72.65)
Mean (SD)	65.9 (13.21)	Female	140 (37.53)	Died	102 (27.34)
IQR	58-75	Total	373	Total	373

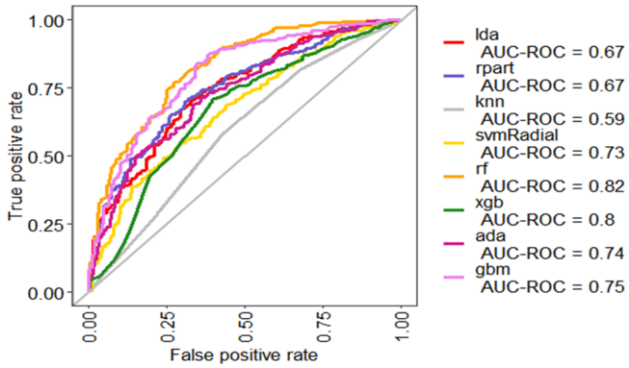


Figure 1. Corresponding AUC-ROC curves of the eight classifiers.

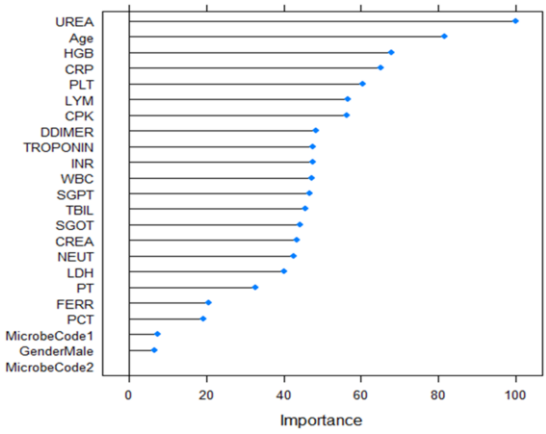


Figure 2. Variable Importance with Random Forest

4. Discussion

Machine Learning approaches are increasingly used to support medical diagnosis or treatment [10, 11]. In this study, we evaluated eight classifiers to predict ICU mortality in critically ill COVID-19 patients. A random forest model outperformed other models in terms of AUC ROC among the various classification methods that were tested using a 10-fold CV. Renal function worsening and older age seem to be among the strongest predictors of unfavorable outcomes in compliance with similar studies [2,4,5,7,12]. Recent literature suggests that predictive modeling can be enhanced by using ML approaches, since they enable the detection of complex relationships between the outcomes of interest and the covariates, especially when trained in the individual data of a hospital, overcoming the limitations of traditional methods [2,3,6,13]. It is difficult to compare the performance of different predictive models, because each one was developed on patients with different characteristics, with different sets of attributes, and with different model development techniques [7].

There are several limitations to this study. First, it is a single-center study with data collected from a single ICU, restricting the generalizability and external validity of the

results. Second, the dataset is limited to demographic characteristics and routine laboratory results retrieved from the hospital information system (HIS) and does not contain clinical patient data, which would undoubtedly add significant accuracy to the predictive ability of the machine learning models. Third, a serious drawback of the proposed ML model (RF), is its limited sensitivity (max 0.5), which is apparently due to the lack of clinical data from the dataset. We anticipate that our future studies besides laboratory and demographic data will additionally include clinical data from patients over longer time periods.

5. Conclusion

We evaluated simple demographics, routine laboratory tests, and the presence of bloodstream infection during the first two days after admission to the ICU, for the prediction of ICU mortality. The best performance regarding AUC-ROC was observed with Random Forest (0.82), with the highest predictive values being urea, age, hemoglobin, CRP, platelet, and lymphocyte count.

References

[1] World Health Organization. *WHO Coronavirus (COVID-19) Dashboard*. https://covid19.who.int/" https://covid19.who.int/ [Accessed 4th March 2023].
[2] Magunia H, Lederer S, Verbuecheln R, Gilot BJ, Koeppen M, Haeberle HA, et al. Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. *Crit Care*. 2021; 25(1):295. doi: 10.1186/s13054-021-03720-4.
[3] Laino ME, Generali E, Tommasini T, Angelotti G, Aghemo A, Desai A, et al. An individualized algorithm to predict mortality in COVID-19 pneumonia: a machine learning based study. *Archives of Medical Science*. 2022;18(3):587-595. doi: 10.5114/aoms/144980.
[4] Jamshidi E, Asgary A, Tavakoli N, Zali A, Setareh S, Esmaily H, et al. Using Machine Learning to Predict Mortality for COVID-19 Patients on Day 0 in the ICU. *Front Digit Health*. 2022; 3:681608. doi: 10.3389/fdgth.2021.681608.
[5] Ustebay S, Sarmis A, Kaya GK, Sujan M. A comparison of machine learning algorithms in predicting COVID-19 prognostics. *Intern Emerg Med*. 2023;18(1):229-239. doi:10.1007/s11739-022-03101-x.
[6] Elhazmi A, Al-Omari A, Sallam H, Mufti HN, Rabie AA, Alshahrani M, et al. Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU. *J Infect Public Health*. 2022;15(7):826-834. doi: 10.1016/j.jiph.2022.06.008.
[7] Lorenzoni G, Sella N, Boscolo A, Azzolina D, Bartolotta P, Pasin L, et al. COVID-19 ICU mortality prediction: a machine learning approach using SuperLearner algorithm. *J Anesth Analg Crit Care*. 2021; 1:3. doi:10.1186/s44158-021-00002-x.
[8] Kuhn M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. 2008; 28(5):1–26. doi:10.18637/jss. v028.i05
[9] R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/. [Accessed 4th March 2023].
[10] Sakagianni A, Feretzakis G, Kalles D, Koufopoulou C, Kaldis V. Setting up an Easy-to-Use Machine Learning Pipeline for Medical Decision Support: A Case Study for COVID-19 Diagnosis Based on Deep Learning with CT scans. *Stud Health Technol Inform*. 2020; 272:13-16. doi: 10.3233/SHTI200481.
[11] Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Martsoukou M, Lada M, et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. *Antibiotics*. 2020;9(2):50. doi: 10.3390/antibiotics9020050.
[12] Afrash MR, Shanbehzadeh M, Kazemi-Arpanahi H. Predicting Risk of Mortality in COVID-19 Hospitalized Patients using Hybrid Machine Learning Algorithms. *J Biomed Phys Eng.* 2022;12(6):611-626. doi: 10.31661/jbpe.v0i0.2105-1334.
[13] Choi MH, Kim D, Choi EJ, Jung YJ, Choi YJ, Cho JH, et al. Mortality prediction of patients in intensive care units using machine learning algorithms based on electronic health records. *Sci Rep.* 2022;12(1):7180. doi: 10.1038/s41598-022-11226-4.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230201

Interdisciplinary Teams in Health Informatics: Using FHIR Standards to Share Computable Knowledge

Elisavet ANDRIKOPOULOU[a,1], Björn SCHREIWEIS[b] and Michael ANYWAR[b]
[a] *University of Portsmouth, Portsmouth, United Kingdom*
[b] *Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein, Campus Kiel, Germany*
ORCiD ID: Elisavet Andrikopoulou https://orcid.org/0000-0002-4614-2907
Michael Anywar https://orcid.org/0000-0001-8028-803X
Björn Schreiweis https://orcid.org/0000-0002-1748-1563

Abstract. The use and shareability of Clinical Quality Language (CQL) artefacts is an important aspect in enabling the exchange and interoperability of clinical data to support both clinical decisions and research in the medical informatics field. This paper, while basing on use cases and synthetic data, developed purposeful CQL reusable libraries to showcase the possibilities of multidisciplinary teams and how CQLs could be best used to support clinical decision making.

Keywords. Clinical Quality Language (CQL), CQL artefacts, Healthcare data, Interoperability, FHIR, computable knowledge

1. Introduction

Creating and sharing computable knowledge, in other words executable knowledge in healthcare is a complex and challenging [1]. To develop computable knowledge, we require data and knowledge, which are formatted in a sequential format, understandable by a computer. As computable knowledge infers that it is not just data, but we require knowledge to produce it. This knowledge can be healthcare based such as causes of a health condition or effectiveness of medication [2] but it also requires engineers and developers, overall organised interdisciplinary teams [3].

The National Institute for Health and Care Excellence (NICE) is trying to increase the level of programmable capabilities of the Clinical Practice Guidelines (CPG), by adding structured and standardised clinical codes to its guidelines [4]. CPGs are meant to express best practices in healthcare and are commonly presented as narrative documents communicating care processes, decision making, and clinical case knowledge. The NICE initiative is clearly related with programming technologies that would enable those codes to be used.

Two of the main strategical objectives of Health Level Seven International (HL7) [5], SNOMED International [6] and Health Data Research UK (HDRUK) [7] are

[1] Corresponding Author: Elisavet ANDRIKOPOLOU, University of Portsmouth, UK.
E-mail: elisavet.andrikopoulou@port.ac.uk

international and national interoperability as well as using data and analytics. To achieve this they provide standards and healthcare domain-specific languages (DSL) such as Fast Healthcare Interoperability Resources (FHIR) [8] and Clinical Quality Language (CQL) [9]. Those are used to describe a CPG, and provide a standardized, shareable, computable artefact that leaves little room for misinterpretation or ambiguity. This is one of the reasons why countries like Germany [10] and the UK are working towards adopting FHIR as the framework for healthcare data interoperability [11].

Computable knowledge artefacts are notoriously un-sharable and often hidden behind corporate walls. Developing libraries of structured, tested, meaningful and sharable computable knowledge artefacts and to enhance the Learning Health System, can improve the benefits of innovation and the decision making of clinicians.

This paper aims to showcase how fit to purpose use cases and synthetic data are used to develop purposeful CQL re-usable libraries.

2. Methods

We developed and verified seven use cases in order to create educational tutorials for computable knowledge, as was described in detail in a previous publication [1].

Four use cases were developed with focus on Type 2 Diabetes (T2D) diagnosis based on Type 2 Diabetes diagnostic NICE guidelines [12]. The other three developed use cases focus on prescribing direct oral anticoagulants (DOAC) in patients with Non-Valvular Atrial Fibrillation (AF) and for the treatment and prevention of venous thromboembolism (VTE) [13]. The T2D use cases were inspired by case studies aimed at training clinicians [14] and the DOAC related use cases were inspired by use cases published previously in the Hospitalist [15]. The intention behind creating these use cases for diagnoses and prescription was to showcase that FHIR standards and CQL are able to be used in both of these categories.

A FHIR server was setup, relevant to the use cases resources were created. This resulted into “Condition” and “Observation” FHIR resources which were coded using SNOMED CT and LOINC code systems [16] respectively, while the UCUM [17] standard was used for the value units.

Based on the Condition and Observation FHIR resources, CQL libraries were created to reflect the use cases and population. The CQLs were then tested using Postman as a REST client application, which successfully returned the desired population results.

3. Results

The developed CQL artefacts, use cases and the rest of the resources of this publication are freely available at the GitHub repository [18]. A healthcare professional verified the use cases and the artefacts for medical accuracy, a programmer developed the FHIR server and the CQL libraries and a health informatician was responsible for the coordination and communication.

The T2D use cases were discussing generic elements of the diagnosis and their rationale is modeled in Figure 1. The use cases about the DOAC prescription’s rationale are modelled in Figure 2. The CQL artefacts were created and tested based on Blaze FHIR server. Figure 3 shows code snippets from use case 2 CQL that was created.

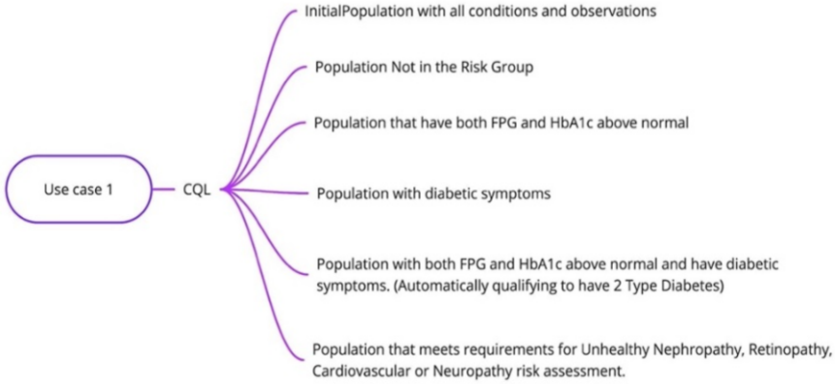


Figure 1. Model of use case 1 (T2D) CQL and population measures that it can query.

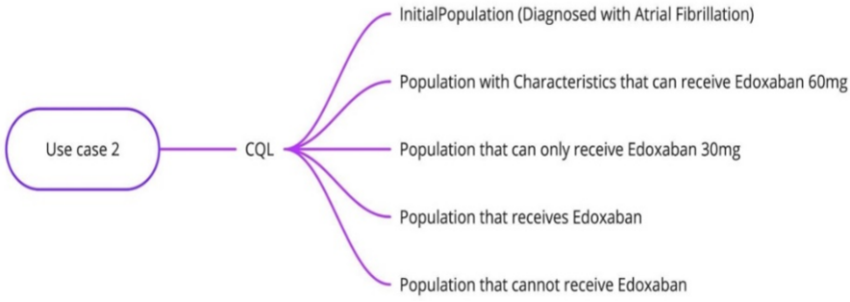


Figure 2. Model of use case 2 (DOAC) CQL and population measures that it can query.

```
//Define Population to exclude from edoxaban 60mg prescription but include in edoxaban 30mg 1 tablet per day prescription
define Thirty_mg_Edoxaban_Prescription:
exists InitialPopulation and
exists ([Observation: Code '35591-7' from LOINC] O where O.value as Quantity >= 30 'mL/min' and O.value as Quantity <=50 'mL/min')

//Define Population to exclude from Edoxaban Prescription
define NO_Edoxaban_Prescription:
not exists "Sixty_mg_Edoxaban_Prescription" or
not exists "Thirty_mg_Edoxaban_Prescription"

//Define Population that can receive Edoxaban.
define Edoxaban_Prescription:
"Thirty_mg_Edoxaban_Prescription" or "Sixty_mg_Edoxaban_Prescription"
```

Figure 3. Model of use case 2 CQL and population measures that it can query.

Further details on the relationship between the use cases and the artefacts can be viewed at our GitHub repository [18]. As a result, reports of type “subject-list” were created. These unlike the “population measure” reports which only list the patient reference id, besides the patient reference ids, of the queried population, they provide detailed information on the patient.

4. Discussion and Conclusion

This paper presents CQL artefacts that work on FHIR servers based on T2D diagnosis and the prescription of DOACs. Working with CQLs requires a Fast Healthcare Interoperability Resources (FHIR) [8] server which supports the CQL modules. Currently not every FHIR servers is capable of this. It is our belief and argued in the literature [19] that the FHIR servers able to support CQL are both too expensive for research and educational teams to afford or they are just not fully supporting all possible CQL queries.

Clinical quality language [9] was initially developed as a quality indicator in US, outside of HL7 standards but soon was incorporated into HL7 only on 2019. Identifying quality indicators evolved into something that can be used for healthcare decision support. In US there are a lot of defined sets of computable knowledge associated with CQL but are not part of it yet, around conditions and procedures that US uses for monitoring. We are working on developing those phenotypes and use them for decision support purposes rather than quality control purposes.

The novelty of CQL presented a problem in our use cases, where we were unable to make references to "*Family Member History*" conditions. This however is a limitation from the FHIR server we used, as it is still in development. We hope that this will be fixed in future revisions, and we will be able to incorporate "*Family Member History*" when diagnosing a patient in our future CQL development.

Overall, the creation of CQL artefacts was a complicated procedure. To correctly diagnose and prescribe, literature on diagnosis and prescription were consulted constantly in order to know the exact conditions under which the described use case phenotypes existed. This highlighted that to correctly model CQL artefacts, interdisciplinary teams are needed of which domain clinicians (medics) need to be part of. Clinical Information Modelling [10] requires highly competent teams that are able to educate all members on interoperability and sharing of resources.

To test and use the CQL artefacts developed for this paper, prior knowledge and the use of SNOMED CT and LOINC code systems are required. This also signifies that developers educated on medical informatics should be part of the interdisciplinary teams when developing CQL artefacts.

4.1. Limitations and future work

Since we did not use all possible code systems for conditions e.g., ICD-10-GM and the different UCUM units for measurements, our artefacts are limited to work with specific code systems and units of measure. This implies that the share-ability of the artefacts besides the complexity in its development is also limited by code systems and unit of measure standards. This however, presents opportunities for future work, which can incorporate all possible code systems in our artefacts to enable cross-border share-ability of CQL artefacts where one does not need to have prior knowledge of the code system the receiver uses to code conditions or observations. This applies to units too especially in cases where an observation value can be stored with more than one unit, for example mg/dL and mg/L.

Overall, we showed that CQL provides a great opportunity for clinical decision support and not just for quality control purposes. And it is through this opportunity that health informaticians and clinicians can be taught how to develop decision support systems based on clinical rules and syntaxes used in CQLs.

Acknowledgements

The authors would like to thank Ms. Jasdeep Jolly from the Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein, Campus Kiel, Germany for her support during the ideation stage and realization of this paper.

References

[1] Andrikopoulou E, Scott P. Experiences of Creating Computable Knowledge Tutorials Using HL7 Clinical Quality Language. Vol. 298, Studies in Health Technology and Informatics. 2022.

[2] Wyatt J, Scott P. Computable knowledge is the enemy of disease. Vol. 27, BMJ health & care informatics. England; 2020.

[3] Scott P. Computable knowledge Part 2 of 2: learning from the HDR UK collaborathon – make it FAIR! [Internet]. 2021 [cited 2022 Jun 30]. Available from: https://www.hdruk.ac.uk/news/computable-knowledge-part-2-of-2-learning-from-the-hdr-uk-collaborathon-make-it-fair/

[4] Mitchell A. A NICE perspective on computable biomedical knowledge. BMJ Heal Care Informatics [Internet]. 2020 Jul 1 [cited 2022 Dec 15];27(2):e100126. Available from: https://informatics.bmj.com/content/27/2/e100126

[5] HL7 International. About Health Level Seven International [Internet]. 2022. [cited 2023 Jan 4]. Available from: https://www.hl7.org/about/index.cfm?ref=nav

[6] NHS Digital. SNOMED CT [Internet]. NHS Digital. 2022 [cited 2022 Jun 30]. Available from: https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct

[7] Health Data Research UK. Our Strategy [Internet]. 2022. [cited 2023 Jan 4]. Available from: https://www.hdruk.ac.uk/about-us/our-strategy/

[8] HL7.org. FHIR Specification [Internet]. [cited 2022 Jun 30]. Available from: https://hl7.org/fhir/

[9] HL7 International. Clinical Quality Language (CQL) [Internet]. [cited 2022 Jun 30]. Available from: https://cql.hl7.org/

[10] Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods Inf Med. 2018 Jul 17;57(S 01):e66–81

[11] Scott PJ, Heitmann KU. Team Competencies and Educational Threshold Concepts for Clinical Information Modelling. Stud Health Technol Inform. 2018;255:252–6.

[12] NICE. When should I suspect type 2 diabetes in an adult? [Internet]. Clinical Knowledge Summaries. 2022 [cited 2022 Jun 30]. Available from: https://cks.nice.org.uk/topics/diabetes-type-2/diagnosis/diagnosis-in-adults/

[13] NHS Greater Glasgow and Clyde Medicines Information Service. DOAC Prescribing in Patients with Non-Valvular AF and for the treatment and prevention of VTE Frequently Asked Questions. Vol. 7. 2018.

[14] Diabetes & Primary Care. Making a diagnosis of type 2 diabetes [Internet]. [cited 2022 Dec 8]. Available from: https://diabetesonthenet.com/cpd-module/making-a-diagnosis-of-type-2-diabetes/

[15] Chadha J, Viquez-Beita K. Can I use DOAC in a patient with renal disease? [Internet]. The Hospitalist. 2022 [cited 2022 Dec 8]. Available from: https://www.the-hospitalist.org/hospitalist/article/31587/renal-genitourinary/can-i-use-doac-in-a-patient-with-renal-disease/

[16] LOINC. Home [Internet]. 2022 [cited 2023 Jan 4]. Available from: https://loinc.org/

[17] National Library of Medicine. Unified Code for Units of Measure. In: Definitions [Internet]. 2020 [cited 2023 Jan 4]. Available from: https://ucum.nlm.nih.gov/

[18] Andrikopoulou E, Anywar M. CzarMich/Kiel-Portsmouth-Collaboration [Internet]. 2023 [cited 2023 Jan 4]. Available from: https://github.com/CzarMich/Kiel-Portsmouth-Collaboration

[19] Li M, Cai H, Zhi Y, Fu Z, Duan H, Lu X. A configurable method for clinical quality measurement through electronic health records based on openEHR and CQL. BMC Med Inform Decis Mak. 2022 Feb;22(1):37.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230202

Extracting Temporal Relationships in EHR: Application to COVID-19 Patients

Carlos MOLINA[a,1] and Belén PRADOS-SUAREZ[a]
[a] *Department of Software Engineering, University of Granada, Spain*

Abstract. Association rules are one of the most used data mining techniques. The first proposals have considered relations over time in different ways, resulting in the so-called *Temporal Association Rules (TAR)*. Although there are some proposals to extract association rules in OLAP systems, to the best of our knowledge, there is no method proposed to extract temporal association rules over multidimensional models in these kinds of systems. In this paper we study the adaptation of TAR to multidimensional structures, identifying the dimension that establishes the number of transactions and how to find time relative correlations between the other dimensions. A new method called *COGtARE* is presented as an extension of a previous approach proposed to reduce the complexity of the resulting set of association rules. The method is tested in application to COVID-19 patients data..

Keywords. Temporal Association Rules (TAR), Multidimensional Model, Complexity, COVID-19

1. Introduction

Association rules are one of the most commonly used methods for decision making. Since the first proposal by Agrawal et al. [1], several extensions of the concept have been proposed. Tung et al. [2] extended the association rules approach considering time relationships between records – what they call *inter-transaction association rules*. The authors proposed an algorithm to extract relations as follows:

when A appears, then B appears T later,

where *A* and *B* are a set of items, and *T* is a measure of time (e.g. *3 days*, *1 month*, etc.). Lu et al. [3] extended the model from one-dimensional inter-transaction (time) to N-dimensional inter-transaction association rules (e.g. time and distance) with application to stock movement prediction. The time concept in association rules has been studied from several perspectives. An exhaustive review and classification can be found in [4].

Association rules have been adapted to other systems that differ from the transactional one. An example are OLAP (On-line Analytical Processing) systems where the data are organized using multidimensional structures called datacubes [5]. One common characteristic of these datacubes is that there is always a dimension to represent time, since OLAPs are used for strategic analysis in organizations. To the best of our knowledge, although there are proposals for association rules over datacubes (e.g. [6,7]), none include time in their methods.

[1] Corresponding Author: Carlos Molina, E-mail: carlosmo@ugr.es.

In this paper we propose a new method to extract temporal association rules from fuzzy datacubes adapted to these analysis-oriented systems. There is no standard for OLAP structures, so we first need to establish the multidimensional model we will use to represent the data. In our case, we use a Fuzzy Multidimensional Model (see [8] for details). As starting point, we will describe an association rules method according to the complexity defined for this structure [7] aimed at getting understandable association rule sets (Section 3.1). The next section presents the datacube applied to COVID-19 patients' data to test the proposed method. The last one presents the main conclusions.

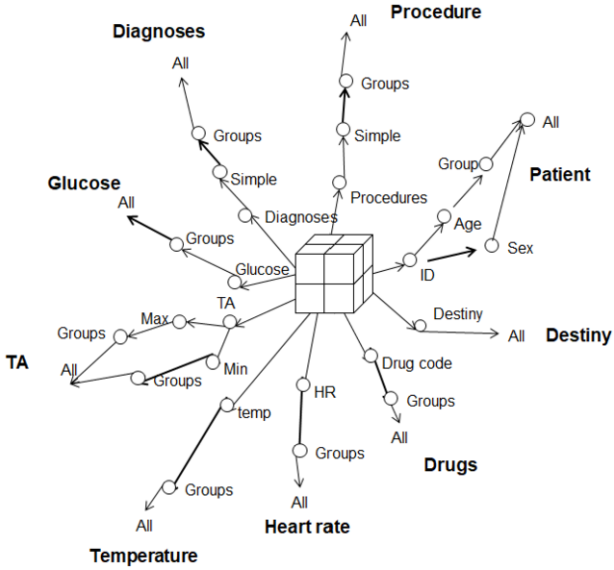


Figure 1. Multidimensional model for COVID-19 patients data (9).

2. Temporal Association Rules over Fuzzy Datacubes

We first need to present a datacube applied to COVID-19 patients' data to test and illustrate the method. Figure 1 shows the multidimensional model. In this example, we have patients' information on COVID-19. The *Patient* dimension includes data of the patients that do not change over time (e.g. age, sex, etc.). In our studies we want to know how the relationships between the other variables evolve over time for each patient, so this dimension will establish the number of records we have (for each patient we will have times series for the rest of the variables and we want to identify temporal relationships common to all the patients). We will refer to these kinds of dimensions whose values do not change over time as *fixed dimensions (D_F)*, which will allow us to identify each of the entities over time.

As we have mentioned before, in most datacubes there is dimension that represents time, which is used for evolution analysis to see trends in the rest of the variables or dimensions. In this example the time dimension is *Time*. This dimension establishes the period in which we have values for the rest of the variables.

The rest of the dimensions have values that change over time for each patient, and we want to find relationships related to time between them. We call these *variable dimensions (D_V)* (e.g. *Drugs* dimension). The last element we need is a way to measure the representativeness and strength of the relationship. For the former, we use the normal measure in association rules: the *Support* (in this example the number of patients that satisfy the rule) and the *Certainty Factor (CF)* [10] instead of *Confidence* because of the known problems with very frequent items [11].

Now we have all the elements to define the *Temporal Association Rules* over datacubes.

Definition. The temporal association rules AR_T are

$$AR_T=(a\cdot b \rightarrow b',t,Support=\alpha,CF=\delta)$$

where:

- $\forall i \in a/i \in D_F$, i are items belonging to the *Fixed Dimensions.*
- $\forall i \in b/i \in D_V$ and $\forall i \in b'/i \in D_V$, $\$$ and i' are items from the *Variable Dimensions*.
- t is a time measure.
- *Support*=α is the representativeness of the rule, where $\alpha \in [0,1]$.
- *CF*=δ is the strength of the relationship, where $\delta \in [-1,1]$.

which means that for a concrete value of time dimension I_T where $a\cdot b$ appears, $a\cdot b'$ appears in the moment I_T+t.

3. Methods

Once we have formally defined the temporal association rules, we present the algorithm to extract them from datacubes. We adapted a previous method (COGARE) that extracts association rules reducing the complexity of the results. The method uses the hierarchies of the dimensions in the datacubes to reduce the number of rules, presenting concepts that are more understandable by users. The method extracts rules over datacubes and diminishes complexity by using the concepts defined in the hierarchy over each dimension to reduce the number of rules and improve interpretability. It comprises three main stages:

- *Itemset generation*: the algorithm uses the Apriori algorithm [1] adapted to the multidimensional model. It takes into account the hierarchy of the dimension when an itemset is not frequent and looks for a generalization that may be frequent (bottom-up approach). In this generalization the support threshold is adapted such that the more general the items are, the higher the threshold is (see [7] for details).
- *Rules generation*: algorithm extracts rules using an Apriori-like approach [1].
- *Rules generalization*: the rule set obtained in the previous stage is generalized to reduce the complexity of the result. In this step, the algorithm tries to generalize the elements in the association rules by defining the items at a higher level (more abstract). If the generalized rules include others (defined over more concrete values but representing the same knowledge) those are deleted. On each step, the quality of the rule set is controlled so if it decreases down to an established threshold, the operation is not applied.

As mentioned before, the associations rules obtained with this process are pruned according to a *certainty factor (CF)* instead of *confidence* to avoid some of the well-known problems of this quality measure.

3.1. COGtARE: COmplexity Guided Temporal Association Rule Extraction

In this section we present the changes made to the COGARE method to extract temporal association rules. As we have mentioned, the concept of support is different. Now we have to calculate the number of different values (records) in the *fixed dimensions*. This can be done querying the datacube applying slice (selecting only the *fixed dimensions*) and applying the values in *variable dimensions* as restrictions.

Once we have the frequent itemsets, next step is the rule generation process. In our approach we have combined two itemsets, one for antecedent and the other for consequent items. But not all itemsets can be combined – they have to be compatible in the sense that both represent the same entity. In our case, the same entity means the itemsets share the same values for the *fixed dimensions*:

Definition. In $I=a\cdot b$ and $I'=a'\cdot b'$ we have two itemsets where 1) $\forall i \in a/i \in D_F$, and $\forall i' \in a'/i' \in D_F$, 2) $\forall i \in b/i \in D_V$, and $\forall i' \in b'/i' \in D_V$, then I and I' are compatible if $a=a'$.

The last aspect is how we calculate the time relation (t) part of the rules. To represent it, we consider the period between the two closest records that satisfy the restrictions (b and b' in the rule) for all the entities that meet the rule. With all these values, the user has a 95% confidence interval. The rest of the steps do not change, so we have all the elements to extract temporal association rules over the datacubes. In next section we test the method over the running examples datacube over COVID-19 patients' data and shows some interesting extracted rules. With this definition, we have executed the proposed algorithm over the datacube in Figure 1. Table 1 shows some examples of extracted rules.

Table 1. Examples of rules obtained.

Rules	Support	CF	t in days (avg and interval)
{}·{Temperature is Normal, Drug N02BE01, Drug V03AN01} → {Destiny is Home}	0.11	0.25	4.3 [3.4,5.2]
{}·{Procedure 0BJ0XZZ applied, Drug D08AC02, Drug V03AN01}→ {Destiny is Home}	0.11	0.33	6.8 [5.6,8.1]
{}·{TA max is high, Drug A02BC01} → {Drug D08AC02}	0.68	0.32	1.3 [1.1,1.5]
{}·{Procedure 0BJ0XZZ applied, Drug V03AN01}→{Destiny is Home}	0.12	0.30	7.2 [5.8,8.5]

4. Conclusions

In this paper we have introduced the time relationships in the extraction of association rules over multidimensional structures in OLAP systems. We have the different roles that dimensions in datacube may play, where we have the ones that identify the entities and which do not change over time (*fixed dimensions*), the dimensions that do change over time, which is where we can find the temporal correlations (*variable dimensions*), and the dimensions that measure the passage of time (*Time dimension*). We have proposed a method to deal with these categories and properly calculate the support of the item sets and the quality of temporal association rules.

In this paper we have applied the proposed method over COVID-19 patient's data.

Acknowledgement

This research is partially supported by B-TIC-744-UGR20 *ADIM: Accesibilidad de Datos para Investigación Médica* of the Junta de Andalucía.

References

[1] Agrawal R, Sritkant R. Fast Algorithms for Mining Association Rules in Large Databases. In: Proceedings of 20th International Conference on Very large data Bases. 1994. p. 478–99.
[2] Tung AKH, Lu H, Han J, Feng L. Breaking the barrier of transactions. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '99. ACM Press; 1999.
[3] Lu H, Han J, Feng L. Stock movement prediction and n-dimensional inter-transaction association rules. In: 1998 ACM SIGMOD Workshop on Research Issues on DMi and KD, Seattle, WA, USA, ACM, New York, USA. 1998.
[4] Segura-Delgado A, Gacto MJ, Alcalá R, Alcalá-Fdez J. Temporal association rule mining: An overview considering the time variable as an integral or implied component. WIREs Data Min Knowl Discov. 2020 Apr;10(4).
[5] Codd EF. Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. E.F. Codd and Associates; 1993.
[6] Zhu H. On-Line Analytical Mining of Association Rules [Internet] [PhD Thesis]. Simon Fraser University; 1998.
[7] Marín N, Molina C, Serrano JM, Vila A. A Complexity guided algorithm for association rule extraction on fuzzy datacubes. IEEE Tras Fuzzy Syst. 2008;16:693–714.
[8] Molina C, Rodriguez-Ariza L, Sanchez D, Vila MA. A New Fuzzy Multidimensional Model. IEEE Trans Fuzzy Syst. 2006 Dec;14(6):897–912.
[9] MH Hospitales. Covid Data Save Lives [Internet]. 2021. Available from: https://www.hmhospitales.com/coronavirus/covid-data-save-lives/english-version
[10] Shortliffe E, Buchanan B. A model of inexact reasoning in medicine. Math Biosci. 1975;23:351–79.
[11] Delgado M, Marin N, Sanchez D, Vila MA. Fuzzy Association Rules: General Model and Applications. IEEE Trans Fuzzy Syst. 2003;11(2):214–25.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230203

The PrescIT Knowledge Graph: Supporting ePrescription to Prevent Adverse Drug Reactions

Achilleas CHYTAS[a,1], Vlasios DIMITRIADIS[a], Giorgos GIANNIOS[b], Margarita GRAMMATIKOPOULOU[b], George NIKOLAIDIS[c], Jenny PLIATSIKA[c], Martha ZACHARIADOU[c], Haralampos KARANIKAS[d], Ioannis KOMPATSIARIS[b], Spiros NIKOLOPOULOS[b] and Pantelis NATSIAVAS[a,1]

[a] *Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece*

[b] *Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece*

[c] *Ergobyte SA, Thessaloniki, Greece*

[d] *Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece*

ORCiD ID: Achilleas CHYTAS https://orcid.org/0000-0001-8486-011X

Abstract. Adverse Drug Reactions (ADRs) are an important public health issue as they can impose significant health and monetary burdens. This paper presents the engineering and use case of a Knowledge Graph, supporting the prevention of ADRs as part of a Clinical Decision Support System (CDSS) developed in the context of the PrescIT project. The presented PrescIT Knowledge Graph is built upon Semantic Web technologies namely the Resource Description Framework (RDF), and integrates widely relevant data sources and ontologies, i.e., DrugBank, SemMedDB, OpenPVSignal Knowledge Graph and DINTO, resulting in a lightweight and self-contained data source for evidence-based ADRs identification.

Keywords. Adverse Drug Reactions, Drug Safety, Clinical Decision Support Systems, ePrescription

1. Introduction

Adverse Drug Reactions (ADRs) have been identified as a major public health issue as they lead to huge healthcare costs and they can also be considered a significant causal factor for health morbidity and mortality [1]. Indicatively, it has recently been quantified that more than 20% of hospitalizations in a random selection of multiple hospital records is related with ADRs.

As Artificial Intelligence (AI) and other relevant technical paradigms have emerged, they have also been widely investigated to support Drug Safety (DS) [2,3]. Focusing on symbolic AI, i.e., the branch of AI which is oriented on "rule-based" knowledge schemes

[1] Corresponding Authors: Achilleas Chytas and Pantelis Natsiavas, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou – Thermi Rd, GR57001, Thessaloniki, Greece; E-mails: achytas@certh.gr, pnatsiavas@certh.gr.

and automatic reasoning upon them, Knowledge Engineering (KE) has been highlighted as a key scientific domain used to support relevant decision support systems [4]. KE includes the use of Natural Language Processing (NLP) to extract knowledge from free-text, the use of ontologies and specific data formalisms to represent knowledge - typically in the form of Knowledge Graphs (KGs) - and the use of reasoning algorithms to infer new knowledge upon the explicit statements stored in the respective knowledge base.

Clinical Decision Support Systems (CDSS) have also been developed and are currently used to support the prevention of potential ADRs during various aspects of the clinical practice (e.g., ePrescription, clinical orders etc.). Typically, CDSSs are integrated in larger healthcare systems like Electronic Health Record (EHR), Computerized Physician Order Entry (CPOE), ePrescription systems etc. The PrescIT project[2] is a nationally funded research and development initiative aiming to develop a CDSS platform to support safe ePrescription via the prevention of ADRs. To this end, a CDSS is developed and will be pilot tested in the clinical context – the consortium includes three clinical partners. PrescIT employs KE as one of its main technical paradigms and the deployment of a KG, as well as other technical components, e.g., a dynamic workflow module using Business Process Management Notation (BPMN) to employ clinically validated Therapeutic Prescription Protocols [5,6].

This paper focuses on the description of the PrescIT KG, the main module used to deploy the rules upon which the CDSS builds its "alerts" to prevent potential ADRs.

2. Methods

A modular architecture is employed for the PrescIT CDSS, according to which each module can be considered a standalone service integrated with other modules via HTTP calls. The main module elaborated in this paper is the PrescIT KG, consisting of several openly available data sources. The core of the system are the 4 major knowledge sources which are represented as KG using OWL/RDF as the main data formalism:

1. SemMedDB: A knowledge base containing information extracted from thousands of PubMed papers via NLP [7].
2. OpenPVSignal KG: More than 100 pharmacovigilance signal reports published by Uppsala Monitoring Centre, using OpenPVSignal as the main ontological model to represent them [8].
3. DrugBank: An up-to-date, widely used and free-to-access online database containing information on a variety of drugs and drug targets. It combines general useful information regarding drugs such as chemical, pharmacological and pharmaceutical data [9].
4. The DINTO ontology: An RDF based knowledge source integrating various data sources containing information about drug-drug interactions [10].

OpenPVSignal KG was created via a manually curated process with various stages of quality control, while the SemMedDB/DrugBank KGs were populated via scripts with a subset of the data retrieved from their respective data sources. This process of converting already existing data in RDF/OWL format, required significant engineering work directly related with each data source's specifics. Indicatively, for SemMedDB the original data are presented in a relational triple-based format (SQL) i.e., subject-predicate-object (s, p, o). In order to extract the desired information, the fields

[2] https://www.prescit.com

"PREDICATE", "SUBJECT_SEMTYPE" and "OBJECT_SEMTYPE" were filtered. For PREDICATE only the "CAUSES" records were selected, for SUBJECT_SEMTYPE all terms that may allude to drugs were selected, whilst for OBJECT_SEMTYPE all terms that may refer to an ADR were selected. In the end, the selected (s, p, o) were equivalent to a pattern logic like similar to "biochemical substance" "causes" "condition", also pointing to the relevant article's PubMed id. It should be noted that the SemMedDB KG also interlinks with other widely used terminologies (e.g., ATC, MedDRA, etc.) via the concept unique identifier, part of the UMLS Metathesaurus.

A simple ontological model, which also has the potential to minimize reasoning execution times, was selected for the KG design since its goal is to be utilized mostly as a data source for ADRs, DDIs and their evidence. The KG is annotated and interlinked with other data sources ontologies, not only unifying the PrescIT KG, but also for providing additional querying options to the end users (i.e., either name of a disease or MedDRA/SNOMED codes etc.). For each such aspect of the KG, the most suitable data sources were selected.

3. Results

The PrescIT KG is hosted in two triple stores, based on Virtuoso and Ontotext GraphDB platform (fig. 1). The data are consumed using a set of SPARQL queries. Indicative queries can be outlined as follows: "List known/suspected ADRs for drug A", "Given drugs A and B, list all known/suspected ADRs" etc. The produced responses include the evidence in which each data source base these claims.

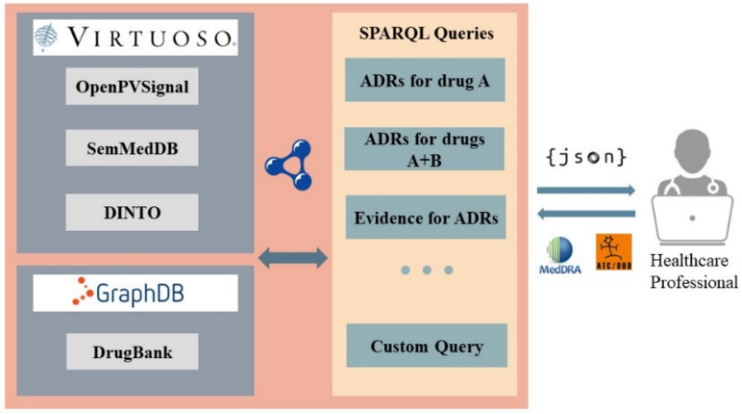


Figure 1. PrescIT architecture

To avoid direct access and enforce technical security controls, in the case of the Virtuoso database, it is exposed via an additional proxy server which was developed using FLASK, a python-based web application framework, while a similar approach is applied for the GraphDB data which are accessible via a REST API. Predefined SPARQL queries were encapsulated and exposed via specific endpoints while an

additional endpoint was created that enabling the submission of custom SPARQL queries. The predefined query endpoints facilitated the data utilization without prior knowledge of the KG "schema" (TBox).

The query results are meant to be consumed by the PrescIT CDSS' front-end where they are consolidated and presented to the end-user, typically a healthcare professional. The total amount of triplets across the entire PrescIT KG is 22.5M (16M triplets from DrugBank, 108k from OpenPVSignal, 5.4M from SemMedDB+UMLS and 1M from DINTO). The KG is available upon request.

4. Discussion

Despite the hype of AI and data science, relevant technical developments have not yet been widely adopted in the clinical context. Hence, there is an increasing need for the development of new tools aiming to integrate "intelligent" technical paradigms to support drug safety.

However, the integration of "intelligent" tools in the clinical context comes with several challenges. Testing, validation and certification of such systems come with ethical, administrative, legal and technical difficulties [11]. To this end, there is an active discussion regarding how AI could be "trustable" in terms of supporting clinical operations. The PrescIT project will soon enter the pilot testing phase, evaluating the impact of the proposed CDSS and the relevant challenges, currently the technical aspects are under validation while the clinical evaluation will take place the following months.

Regarding the implementation and integration of the PrescIT KG, these challenges can be summarized as follows:

- Testing, validation and integration in the clinical environment: Clinical environments are complex and difficult to be standardized as flows of information could heavily vary, even in the same hospital. Three clinical partners have been involved during the KG development to support the definition of the relevant data sources and the PrescIT CDSS design and evaluation as a whole. For an effective integration, the PrescIT CDSS provides various flexible levels of integration with local EHRs via vendor "neutral" interfaces (e.g., SPARQL and REST APIs). Beyond the technical integration difficulties, challenges such as KG's usability and information comprehension, i.e., how well are the "alerts" received etc., due to various issues (e.g., cultural, language barriers etc.) are going to be evaluated during the pilot phase.
- Reasoning and data volume: A technical challenge directly related with the use of the KG, has to with the ability to use a "reasoner", i.e., software which is able to infer RDF statements beyond the ones which are explicitly stated in the KG. This reasoning process is computationally intensive and it has been identified as significant performance bottleneck. In order to avoid this, it was decided that no "reasoner" will be used and the "intelligence" required (e.g., the need to identify relevant RDF individuals based on subclass relationships) will be integrated in the respective SPARQL queries.

5. Conclusion

The PrescIT project aims to support ePrescription process via an integrated CDSS facilitating the prevention of potential ADRs. The PrescIT CDSS is based on the use of an RDF-based KG and its integration in the clinical environment comes with several challenges, also discussed in the paper while the impact of the proposed system and potential gaps are going to be evaluated during the pilot phase. As part of the future work there are plans of extending the KG with biochemical and pathway information which might be clinically relevant. More specifically, pharmacogenomics has been identified as a potential use case for the PrescIT KG.

Acknowledgements

This research has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T2EDK-00640).

References

[1] Formica D, Sultana J, Cutroneo PM, Lucchesi S, Angelica R, Crisafulli S, et al. The economic burden of preventable adverse drug reactions: a systematic review of observational studies. Expert Opin Drug Saf. 2018 Jul;17(7):681–95.
[2] Bates DW, Levine DM, Salmasian H, Syrowatka A, Shahian DM, Lipsitz S, et al. The Safety of Inpatient Health Care. The New England Journal of Medicine. 2023 Jan 11;388(2):142–53.
[3] Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. Trends in Pharmacological Sciences. 2019 Sep 1;40(9):624–35.
[4] Natsiavas P, Malousi A, Bousquet C, Jaulent MC, Koutkias V. Computational Advances in Drug Safety: Systematic and Mapping Review of Knowledge Engineering Based Approaches. Frontiers in Pharmacology. 2019 May 17;10:415.
[5] Natsiavas P, Stavropoulos TG, Pliatsios A, Karanikas H, Gavriilidis GI, Dimitriadis VK, et al. Using Business Process Management Notation to Model Therapeutic Prescription Protocols: The PrescIT Approach. Public Health and Informatics: Proceedings of MIE 2021. 2021 Jul 1;1089–90.
[6] Karanikas H, Papadakis M, Thireos E. Development Of Prescription E-Protocols For Medicines And Integration On The Greek National E-Prescription System. Value in Health. 2015 Nov 1;18(7):A385.
[7] Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindflesch TC. SemMedDB: a PubMed-scale repository of biomedical semantic predications. Bioinformatics. 2012 Dec 1;28(23):3158–60.
[8] Natsiavas P, Boyce RD, Jaulent MC, Koutkias V. OpenPVSignal: Advancing Information Search, Sharing and Reuse on Pharmacovigilance Signals via FAIR Principles and Semantic Web Technologies. Frontiers in Pharmacology [Internet]. 2018 [cited 2023 Jan 13];9. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2018.00609
[9] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research. 2006 Jan 1;34(suppl_1):D668–72.
[10] Herrero-Zazo M, Segura-Bedmar I, Hastings J, Martínez P. DINTO: Using OWL Ontologies and SWRL Rules to Infer Drug–Drug Interactions and Their Mechanisms. J Chem Inf Model. 2015 Aug 24;55(8):1698–707.
[11] Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. npj Digital Medicine. 2020 Dec 21;3(1):107.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230204

Interdisciplinary Human-Centered AI for Hospital Readmission Prediction of Heart Failure Patients

Amira SOLIMAN[a,1], Monika NAIR[a], Marcus PETERSSON[b], Lina LUNDGREN[a], Petra DRYSELIUS[b], Ebba FOGELBERG[b], Omar HAMED[a], Kobra ETMINANI[a] and Jens NYGREN[a]
[a] *Halmstad University, Halmstad, Sweden*
[b] *Cambio Healthcare Systems AB, Sweden*

Abstract. The evolution of clinical decision support (CDS) tools has been improved by usage of new technologies, yet there is an increased need to develop user-friendly, evidence-based, and expert-curated CDS solutions. In this paper, we show with a use-case how interdisciplinary expertise can be combined to develop CDS tool for hospital readmission prediction of heart failure patients. We also discuss how to make the tool integrated in clinical workflow by understanding end-user needs and have clinicians-in-the-loop during the different development stages.

Keywords. Interdisciplinary Healthcare, Human-Centered AI, Hospital Readmission Prediction.

1. Introduction

Unscheduled rehospitalization of heart failure (HF) patients has received a lot of research attention due to its high cost and undesirable impact on patient outcomes and healthcare planning [1]. Providing a care plan with rapid follow-up and treatment compliance can be effective in minimizing the risk of readmissions [2], however, identifying which patients to prioritize remains to be a challenge. A wide range of predictive modeling methods have been used to address this problem, however, none have been implemented as clinical decision support (CDS) applications in practice using Artificial Intelligence (AI) models that learn and adapt to patterns expressed in patient data.

Sensitive and eXplainable Artificial Intelligence (XAI) models can assist clinicians in managing HF patients at discharge by highlighting individual aspects associated with high-risk of readmission. Previous studies showed the potential applicability in terms of cost savings [3]. However, to gain trust by clinicians, interpretability of model results is needed. Additionally, from a human-centered design perspective, interpretability is not a property of the Machine Learning (ML) model only, but a relationship between the model and the user [4]. Human-centered AI systems operating jointly, rather than alone, have a great potential for high effectiveness [5], however, the design process is complex and requires interdisciplinary as well as interorganizational collaborations.

[1] Corresponding Author: Amira Soliman, E-mail: amira.soliman@hh.se.

The objective of this paper is to present a use case developing a CDS application to predict HF patient risk of readmission. The work process had the following parts: 1) identify barriers and enablers for implementation of a CDS application, 2) develop an XAI model, and 3) explore how to present the model output to users.

2. Methods

This research was composed of three interlinked work packages (WP) as shown in Figure 1. The group consisted of university researchers from computer science, UX-design and implementation science, one IT company with expertise in data science and UX-design, and a healthcare organization providing care to HF patients.

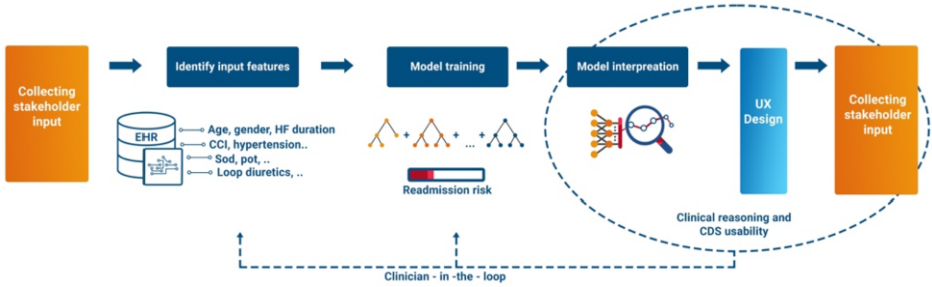


Figure 1. Adopted framework for CDS development, including collection of stakeholder needs, model calibration, clinical reasoning of model/XAI outcome, and checking usability of the developed prototype.

2.1. Collecting Stakeholder Input

To ensure relevance of the application to the clinical context, input from stakeholders in Region Halland in Sweden was assembled [6]. Twelve interviews were performed in 2021 with medical specialists in cardiology, specialist nurses, a physiotherapist, a home care physician, a home care nurse and a controller. The semi-structured interviews were conducted one-to-one via video calls, recorded and transcribed verbatim. The questions concerned potential influence the application might have on patients, healthcare staff, and the organization, as well as what could be the barriers and enablers of potential implementation of such a system in practice. The data were thematically coded independently by two researchers [7] using the categories of the NASSS implementation framework [8].

2.2. AI-enabled Readmission Prediction

Retrospective data from the regional healthcare information platform in Region Halland was used for model development [9]. The cohort consisted of patients ≥40 years of age, diagnosed with HF according to ICD-10 (I11.0, I42, I43, and I50), who had at least a single admission after being diagnosed with HF between January 1, 2017, and December 31, 2019. An ML model was developed with CatBoost resembles gradient-boosting decision trees that handles categorical and continuous features [10]. The model makes predictions using a series of decision trees and was trained using k-fold cross-validation for 10 iterations using the training dataset. The SHapley Additive exPlanations (SHAP) technique was adopted to provide more details behind the model decision regarding important features for readmission prediction [11]. SHAP computes a score for each individual prediction to illustrate features that are positively or negatively driving toward

readmission risk. The provided explanations were addessed by phycisians for clinical relevance.

2.3. Prototype Design

Based on stakeholder input, low-fidelity sketches were made and tested with three users. After iterating the design, a high-fidelity prototype was tested with three new users. For each usability test, the think-aloud protocol was used followed by a short discussion. All users were practicing or former clinicians with HF expertise.

3. Results

Overall, the reasons for iterations between WPs were; need for additional input data to the model (information-based decisions), stakeholder input on when and how to use the tool (process-based decisions), and stakeholder input on information relevance (interface and utility-based decisions).

3.1. Stakeholder Input

Reducing subjectivity in assessing the risk of readmission for patients was seen as a key value for the clinical context. The stakeholders urged that units outside of cardiology dealing with HF patients, mainly due to a lack of beds, should have access to the CDS application to direct the right patients to the cardiology unit and ensure the right treatment and a reduced risk of future readmission. Free-text medical notes were deemed highly important for the algorithm, as they hold relevant information. The risk assessment should be available early, during admission, thereby enabling contacts with cardiology unit, optimal resource allocation, and prioritization. In addition to risk assessment, the application should present factors that led to such assessments. Lack of interoperability with the existing IT applications in the hospital could hinder adoption. A pilot study, information campaign, training, resource allocation, revised procedures, and competent implementation leaders persons were mentioned as powerful enablers for successful adoption of the application.

3.2. Model Performance and Explainability

The model produced a probability between 0 and 1 to indicate the likelihood of readmission. We used a threshold of value (0.43) to convert the generated probability to classification label, i.e, to separate patients with high readmission risk from others. The threshold value was estimated during model training. Thus, a probability below 0.43 indicates low-risk with label 0, while a probability above 0.43 indicates high-risk of readmission with label 1. Table 1 displays the achieved performance by the model on unseen testing data. Interestingly, the developed model gained better performance compared to both LACE index and HOSPITAL score that are widely used as risk assessment tools to discriminate readmissions using readily available clinical predictors that can be calculated before discharge, yet they are not trainable as our model.

Model explainability is key to clarify the risk. Feature ranking generated by SHAP shows that besides clinically informative variables, administrative features have high importance, e.g., patients with frequent inpatient episodes (Figure 2a). While these features cannot be altered, they are useful warning signals when planning for the discharge.

Table 1. Results obtained from training, validation, and testing model with data from Region Halland

	CatBoost Model	LACE Index	HOSPITAL Score
Sensitivity	80	40	49
Specificity	51	77	67
F1-Score	43	35	35
AUC	55	59	58
AUCPR	57	42	43

Abbreviations: AUC, area under receiver operating characteristic curve; AUPRC, area under the precision-recall curve (AUPRC). The reason behind lower specificity ratio is due to sensitivity-based training adopted to be relevant in clinical operations; a lower specificity (minimum of 50%) is accepted to reach higher sensitivity.

3.3. Prototype Design

The findings of the usability tests with clinicians resulted in an interface divided into four parts (Figure 2b). We identified a need to adapt the model score to the right context to be understandable. Presenting the binary output (e.g. high risk vs low risk) together with the model score resulted in some misinterpretations, although both might still be needed for the information to be understandable. In one design iteration we also included information on the model confidence (i.e. how often the model was correct on the training data for a given model score interval), but this was perceived as too complex by the users. Further, we noticed that additional information was needed for the users to trust and use the SHAP values, e.g. previous lab values and admission reasons. We also recognized that parameters that were converted into multiple binary features, such as normal and abnormal, would be more understandable if grouped together in the user interface. Finally, the language we use and what information we choose to present are important for the user's overall understanding of and trust in the result.

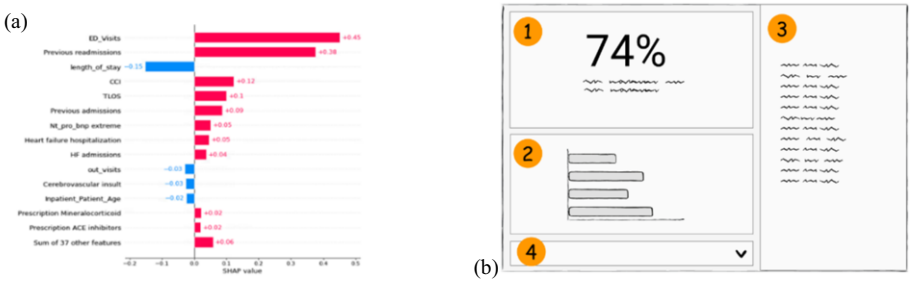


Figure 2. From (a) model prediction of a correct readmission case with 74% probability and list of important features from SHAP to (b) a low-fidelity sketch of the user interface. (a) Color indicates which features increase (red) or decrease (blue) the readmission risk, value on x-axis reflects strength of impact on prediction, features are sorted in y-axis according to their importance. (b) 1) Main information: Risk assessment and potentially in-depth explanation of meaning of model output and confidence. 2) SHAP values: Explanation followed by features with top 5 pos/neg SHAP values. Possible to access information about other contributing features. 3) Details about each feature (history and definitions). 4) Access to general model information and possibly links to additional information about model and results.

4. Discussion and Conclusion

This paper brings up a use case of working with human-centered AI for healthcare. This type of systems (i.e. CDS applications) are intended to improve healthcare delivery by enhancing medical decisions with knowledge extracted from patient data, targeted clinical knowledge, and other health information. Development of such tools requires the skills of multiple experts across different domains, i.e., medical, artificial intelligence,

implementation science, and UX design, however, also requires cross-disciplinary understanding to translate research outputs into new information [5].

The development process, although agile and adaptive, might face some challenges due to uncertainty caused by the complexity of the healthcare context [12]. One such example was related to what point in the care process to place the CDS application, an aspect that needs definition already in the model development process. Project progress showed several alternatives, due to the organisation of the care process and specific context, requiring further iterations with stakeholders, in combination with objective evaluation of the AI-model performance. This, however, is still in silico, as we cannot clinically test a solution at its infant stages.

Other challenges were related to model output presentation, such that the tool and the XAI can be understood and provide the value sought after by users, as well as continuous dialogues with clinical professionals. Expert knowledge integration into the project was a precondition to develop and design both the AI-model and the interface prototype, thus required careful planning to make the most out of the limited time together with clinical professionals.

In conclusion, the project showed that developing CDS applications are feasible and embraced by potential users. However, there are uncertainties and challenges one should carefully consider and plan for in this type of project.

Acknowledgment and Ethical considerations

This project is part of the CAISR Health research profile funded by the Knowledge foundation. The data analysis is approved by the Swedish Ethics Review Authority (Etikprövningsmyndigheten) 2016/517 with Dr. Markus Lingman as chairperson.

References

[1] Yasin ZM, Anderson PD, Lingman M, et al. Receiving care according to national heart failure guidelines is associated with lower total costs - an observational study in Region Halland, Sweden. Eur Heart J Qual Care Clin Outcomes. 2021;7:280-286.

[2] Wideqvist M, Cui X, Magnusson C, Schaufelberger M, Fu M. Hospital readmissions of patients with heart failure from real world: timing and associated risk factors. ESC Heart Fail. 2021;8:1388-1397.

[3] Liu CF, Chen ZC, Kuo SC, Lin TC. Does AI explainability affect physicians' intention to use AI? Int J Med Inf. 2022;168:104884. doi:10.1016/j.ijmedinf.2022.104884

[4] Chen H, Gomez C, Huang CM, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. NPJ Digital Med. 2022 Oct 19;5(1):1-5.

[5] Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 7 (2019): e7702.

[6] Franco-Trigo, L., et al. Stakeholder analysis in health innovation planning processes: a systematic scoping review. Health Policy 2020;10(124) 1083-1099.

[7] Saldaña, J. The coding manual for qualitative researchers. London, UK: SAGE. 2015

[8] Greenhalgh, T., et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC medicine, 2018;16(1) 1-15.

[9] Ashfaq A, Lönn S, Nilsson H, et al. Data Resource Profile: Regional healthcare information platform in Halland, Sweden. Int J Epidemiol. 2020;49:738-739.

[10] Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems 31; 2018.

[11] Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020;2(1):56-67.

[12] Tessarolo F, Nollo G, Conotter V, et al. User-centered co-design and AGILE methodology for developing ambient assisting technologies: Study plan and methodological framework of the CAPTAIN project. In: IEEE 23rd International Symposium on Consumer Technologies (ISCT) 2019 Jun 19. 283-286.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230205

Patient Electronic Health Record as Temporal Graphs for Health Monitoring

Hugo LE BAHER[a,b,c,1], Jérôme AZÉ[b], Sandra BRINGAY[a,d], Pascal PONCELET[a], Nancy RODRIGUEZ[a] and Caroline DUNOYER[c,e]
[a] *LIRMM, UMR 5506, Université de Montpellier, CNRS, Montpellier, France*
[b] *5 DEGRÉS, Paris, France*
[c] *Département d'Information Médicale, CHU Montpellier, Montpellier, France*
[d] *AMIS, Université Paul-Valéry, Montpellier, France*
[e] *IDESP, UMR UA11, INSERM - Université de Montpellier, Montpellier, France*

Abstract. Machine learning methods are becoming increasingly popular to anticipate critical risks in patients under surveillance reducing the burden on caregivers. In this paper, we propose an original modeling that benefits of recent developments in Graph Convolutional Networks: a patient's journey is seen as a graph, where each node is an event and temporal proximities are represented by weighted directed edges. We evaluated this model to predict death at 24 hours on a real dataset and successfully compared our results with the state of the art.

Keywords. Graph neural networks, MIMIC-III, health monitoring, mortality prediction, electronic health record

1. Introduction

Supervision of patient status in medical emergency services is a difficult problem. Jung et al. [1] show that the implementation of decision support tools, which are able to anticipate critical risks in patients under surveillance, would improve the capacity of services to react as fast as possible to unexpected episodes and thus preventing the deaths of their patients. Different approaches have been proposed to predict medical outcomes of a patient from his history. Typically, all data are transformed into vectors in order to use traditional learning techniques such as partitioning, k-nearest neighbor [2], logistic regression, or deep learning models [3]. Under the assumption that the variations over time of the patient's descriptors are essential for predicting his future state, other techniques consider a patient journey as a sequence, consisting of a succession of vectors. This representation allows the use of recursive models like LSTM [4]. The implementation of these techniques requires slicing the patient's history, according to time. Finding the cut protocol that best represents the whole data captured over time is an expensive process and the success of this research influences the predictive ability of the model. For example, if we consider the granularity of the hour, we lose the evolution of the heart rate second by second. On the contrary, if we consider a breakdown to the second, the representation produced is too voluminous to be inputted into most models.

[1] Corresponding Author: Hugo LE BAHER, E-mail: hugo.le-baher@lirmm.fr

Another underlying problem with this representation is that the different patient sequences have varying lengths which require adaptations in the learning model. Recently, GNN (*Graph Neural Network*) have shown to be very suitable for making predictions [5]. Considering the aforementioned objectives and the identified limits of the approaches in the literature, we propose in this article a modeling of patient records as temporal graphs, input into GNNs.

The contributions of the article are as follows: 1) We propose a new original modeling of temporal and heterogeneous data in the form of a graph where each node corresponds to a measurement and where temporal proximities are represented by weighted arcs. 2) We input this representation into a GCN (*Graph Convolutional Network*) that we evaluate on the MIMIC-III dataset described by [6], to predict the patient death at 24 hours, from real-world data collected during their journeys. We show through experiment that our approach offers results comparable to those of the state of the art.

The rest of the article is organized as follows. Section 2 describes the modeling proposed to directly integrate the time dimension. Section 3 describes the experiments carried out. Finally, we conclude by presenting our future work in Section 4.

2. Method

We propose a modeling of data from patient files in the form of graphs which aims to avoid a temporal slicing. Formally, records of a given patient are represented as a graph $G = (E, P, X)$ where $E = \{e_1, e_2, \ldots, e_n\}$ represents the events, $P = \{(e_i, e_j, p)\}, 1 \leq i \leq n, 1 \leq j \leq n, p \in \mathbb{R}^+$ represents the set of weighted arcs with the temporal proximity p between two events e_i and e_j and $X \in \mathbb{R}^{n \times m}$ represents the characteristics of events with m number of distinct categories measured. Our objective is to predict the medical outcome of interest, i.e., death at 24 hours. This is a binary classification problem, namely finding a mapping f such that $f: G \rightarrow, Y \in \{0,1\}$, Y matches the binary label.

The particularity of the modeling resides in the constitution of the arcs. Information propagates through the convolution process within the final model. The proposed hypothesis is that events close in time have comparable properties. This translates into the graph by calculating a temporal proximity p for each couple (e_i, e_j) from their temporal coordinates $T = \{t_1, t_2, \ldots, t_n\}$ which acts as a weight for the arcs. Thus, the closer two events are in time, the greater the weight of the arc that connects them. The gradual decrease of the weight as the temporal distance increases calls for the study of forgetting from cognitive sciences. We compared several forgetting functions by experimentation and retained the function developed in [7] Π , which is defined by:

$$\Pi(b, \delta_{i,j}) = \exp(-b \cdot \delta_{i,j}) * (\delta_{i,j} \geq 0), b > 0, \delta_{i,j} = t_i - t_j$$

δ corresponds to the temporal distance computed for each pair of events. b is a modeling parameter, which influences the intensity of the decrease in the function. The value of this parameter has a direct effect on the structure of the generated graphs.

Figure 1 summarizes the different stages of the proposed modeling, which starts from the data to produce a temporal proximity graph: on the one hand, the centered value and the modality of each measurement are transformed into a matrix where each column

corresponds to a measured modality **(1)**. Each line corresponds to the vector representation of a given node. On the other hand, the temporal coordinates of events are transformed pairwise into distances **(2)**. The distances are then transformed into proximity by the application of the function Π **(3)**. The result is an adjacency matrix that describes the weight of the edges of the final graph. The adjacency matrix and node representation vector put together describe the graph of a patient journey **(4)**, inputted in a supervised model to predict the binary label, graph-wise.

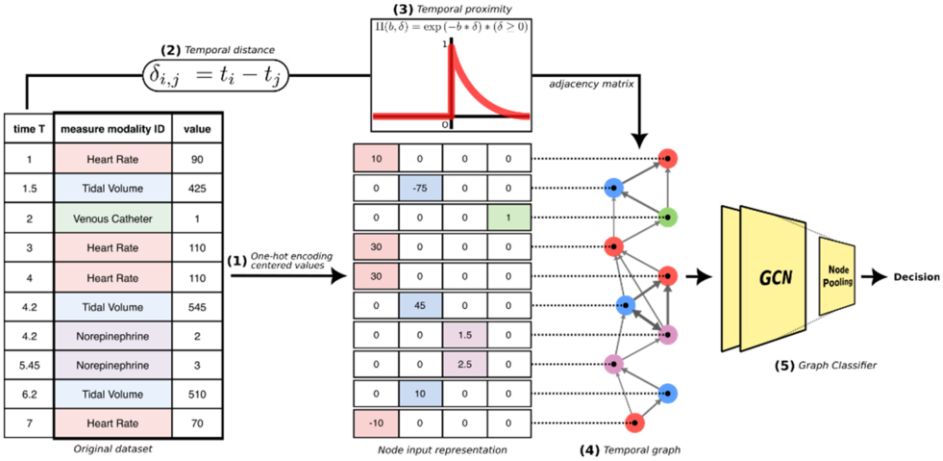


Figure 1. Transformation process of the original data into a temporal graph of measurements. Illustration based on fictional data. **(1)** Transformation of the values through centered one-hot encoding. **(2)** Pairwise distance of temporal coordinates. **(3)** Computation of the temporal proximity. **(4)** Patient representation: Node features and adjacency matrix which describe the graph. **(5)** Graph classification model.

3. Experiments and Results

To evaluate our approach, we focused on the 24-hour death prediction task described in Harutyunyan et al. [6]. Only patient journeys that last at least 4 hours are selected. For each of them, we seek to predict for each one-hour span whether a patient dies within the next 24 hours, based on data prior to the observation. The original data is available on Physionet [8] in the public MIMIC-III Clinical Database [9, 10]. This database is often used in the literature to compare approaches for many medical supervision tasks because it is voluminous, heterogeneous and anonymized. Additionally, the data made available describe actual patient journeys that occurred during Beth Israel Deaconess Medical Center activity from 2001 to 2021. For their experiments, Harutyunyan et al. [6] select 17 descriptors (e.g., Heart Rate, Blood Pressure) according to medical criteria. Values are filtered, pre-processed and ordered categorical data are transformed into continuous. Data volumes are described in Table 1. To obtain a fair comparison with our own experiments, we will rely exactly on these transformed data, which are publicly available.

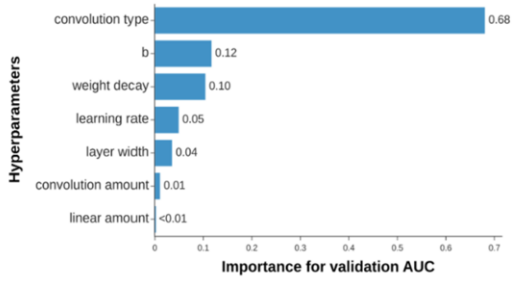
Table 1. Characteristics of data extracted from MIMIC-III.

Amount of patients in training set	28 620
Amount of patients in test set	5 058
Percentage of patients from the minority class	11.0%
Average length of stay	86.6h
Maximum stay duration	2 103h

The model presented is chosen after the search for optimal hyperparameters and structure thanks to the Optuna library [11]. The results presented through the application of our method come from a cross-validation with 5 folds, sliced in the training set then evaluated on the test set. The model referred as "Convolution of Graphs" in the Figure 2 consists of a succession of the following layers: Three layers of GraphSAGE [12] of the DGL library (DeepGraph Library) with the average as aggregation function, then the concatenation of three aggregation functions *min*, *max* and *average* on the representation of all nodes, two linear layers and a *softmax* output. Each trained layer is followed by an activation function *tanh*, except the last. The chosen optimizer is *Adam* with the learning rate set to 0.003 and the weight decay to 0.05. The temporal proximity parameter is fixed at 4.5 and the width of the successive layers is fixed at 15. Learning is computed on a machine equipped with 40 computing cores, 126GB of RAM and 4 NVIDIA GeForce GTX 1080 graphics cards, each with 11GB of VRAM.

Figure 2. On the left, predictive 24-hour mortality results are shown. The different versions of the LSTM are extracted from the original paper corresponding respectively to "S": Standard, "C": Channel-wise and "DS": Deep Supervision. On the right, the importance of hyperparameters on performance following optimization by Optuna for the Convolution of Graphs is shown

Model	AUC
Harutyunyan et al. [6]:	
Logistic Regression	0.870
LSTM - S	0.892
LSTM - C	0.906
LSTM - C + DS	0.911
Our method:	
Convolution of Graphs	0.897



To compare the approaches, we measure the area under the sensitivity/specificity curve, referred to as *AUC*, on the test set. This score is well suited for datasets with unbalanced label proportions. As seen in the Figure 2, our modeling offers results similar to those of the state of the art, observed in Harutyunyan et al. [6], namely logistic regression and the different variants of LSTM. Our objective in this article is to validate the representation of patient data in the form of a graph. The result obtained by our method exceeds some versions of LSTM.

It is questionable to what extent the predictions of the presented model are influenced by the propagation of information from the GNNs through the arcs resulting from the modeling as expected rather than by other effects. The Figure 2 shows the contributing value of different hyperparameters during Optuna's hyperparameter optimization. The results come from a functional analysis of variance, i.e., fANOVA

[13]. We observe on these results that the most important hyperparameters are the type of convolution and the parameter *b* of the modeling. We deduce that the constitution of the arcs and the way in which the information is propagated within the graph in the model have a strong impact on the measured performances.

4. Conclusions

In this article, we have proposed a new modeling of patient data in the form of a graph. This alternative representation of the data enables to consider time. We use weighted arcs to represent the temporal proximity between two events. We experimented this modeling by taking advantage of graph convolutions. Nodes characteristics are propagated along arcs. The main limitation is the studied population (ICU) and task (mortality). Towards generalization, other population and tasks have to be studied. Considering classification accuracy, our results approach those of the state of the art. Other approaches based on node-level prediction should be considered. Moreover, due to the modularity offered by graph structures and the many recent contributions in the field of GNNs, it is possible to imagine various areas to improve explainability. This project is funded by a CIFRE grant, funded by 5 DEGRÉS, established in collaboration with LIRMM and CHU Montpellier.

References

[1] Jung B, Daurat A, De Jong A, Chanques G, Mahul M, Monnin M, et al. Rapid response team and hospital mortality in hospitalized patients. Intensive Care Med. 2016 Apr;42(4):494-504
[2] Poongodi T, Sumathi D, Suresh P, Balusamy B. Deep Learning Techniques for Electronic Health Record (EHR) Analysis. In: Bhoi AK, Mallick PK, Liu CM, Balas VE, editors. Bio-inspired Neurocomputing. Studies in Computational Intelligence. Springer; 2021. p. 73-103.
[3] El-Rashidy N, El-Sappagh S, Abuhmed T, Abdelrazek S, El-Bakry HM. Intensive Care Unit MortalityPrediction: An Improved Patient-Specific Stacking Ensemble Model. IEEE Access. 2020;8:133541-64.
[4] Ashfaq A, Sant'Anna A, Lingman M, Nowaczyk S. Readmission prediction using deep learning on electronic health records. Journal of Biomedical Informatics. 2019 Sep;97:103256.
[5] Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. AI Open. 2020;1:57-81.
[6] Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Scientific Data. 2019 Dec;6(1):96.
[7] White KG. Forgetting functions. Animal Learning & Behavior. 2001 Aug;29(3):193-207.
[8] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000 Jun;101(23):E215-20. Place: United States.
[9] Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific Data. 2016 May;3:160035.
[10] Johnson A, Pollard T, Mark R. MIMIC-III Clinical Database. PhysioNet; 2016. Version Number: 1.4. Type: dataset. Available from: https://physionet.org/content/mimiciii/1.4/.
[11] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2019.
[12] Hamilton W, Ying Z, Leskovec J. Inductive Representation Learning on Large Graphs. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc.; 2017.
[13] Hutter F, Hoos H, Leyton-Brown K. An Efficient Approach for Assessing Hyperparameter Importance. In: Proceedings of International Conference on Machine Learning 2014 (ICML 2014); 2014.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230206

Feature Selection Based on a Genetic Algorithm for Optimizing Weaning Success

Samanta ROSATI[a], Andrea SCOTTO[a], Vito FANELLI[b] and Gabriella BALESTRA[a,1]
[a] *Department of Electronics and Telecommunications, Politecnico di Torino, Italy*
[b] *Department of Surgical Sciences, Università degli Studi di Torino, Italy*
ORCiD ID: Samanta ROSATI https://orcid.org/0000-0003-0620-594X
ORCiD ID: Andrea SCOTTO https://orcid.org/0000-0002-3317-3843
ORCiD ID: Vito FANELLI https://orcid.org/0000-0002-1647-2411
ORCiD ID: Gabriella BALESTRA https://orcid.org/0000-0003-2717-648X

Abstract. Finding the right time for weaning from ventilator is a difficult clinical decision. Several systems based on machine or deep learning are reported in literature. However, the results of these applications are not completely satisfactory and may be improved. An important aspect is represented by the features used as input of these systems. In this paper we present the results of the application of genetic algorithms to perform feature selection on a dataset containing 13688 patients under mechanical ventilation characterizing by 58 variables, extracted from the MIMIC III database. The results show that all features are important, but four of them are essential: 'Sedation_days', 'Mean_Airway_Pressure', 'PaO2', and 'Chloride'. This is only the initial step to obtain a tool to be added to the other clinical indices for minimize the risk of extubation failure.

Keywords. MIMIC, mechanical ventilation, weaning, genetic algorithms, feature selection

1. Introduction

Weaning from mechanical ventilation is the operation of liberating patients from respiratory support that ends with removing the endotracheal tube that is used to attach a patient to a ventilator in intensive care unit (ICU). Finding the right time for extubating a patient is a difficult clinical decision. Delayed weaning is associated with higher risk of patients' muscle atrophy, of nosocomial infections and with ICU's costs. Moreover, extubation failure increases the risk of patients' mortality.

Different support tools are available to estimate the probability of extubation failure (EF). The Rapid Shallow Breathing Index (RSBI) [1] is one of the most used indices for this purpose. However, it has high sensitivity but lower specificity. Different Machine Learning (ML) methods have been used by researchers to provide accurate EF estimation [2–5]. They have good results, but improvement is still needed.

The aim of our work is present some preliminary results related to the construction of a ML-based classifier to be used for the prediction of the weaning outcome. In particular, we focalize the attention on the selection of the most informative features for prediction purposes.

[1] Corresponding Author: Gabriella Balestra, E-mail: gabriella.balestra@polito.it.

2. Dataset Construction

2.1. Feature extraction

The dataset used in this study was extracted from the MIMIC III database [6]. First, from the available data we identified only patients that were supported by a ventilator during their stay in ICU. The number of identified patients was 23968. Then, four exclusion criteria were applied: age less than 18 years, death before extubation, duration of the ventilation lower than 24 hours, unintentional extubation. The result was an initial dataset containing 13691 patients.

For each patient we extracted a total of 91 variables describing personal (age, sex, weight, weight loss, height), and clinical information (SOFA score, GCS score, comorbidities, minimum, maximum and mean values of physiological parameters, laboratory parameters), weaning parameters, and the administered therapy. Patients were assigned to two classes: patients that were extubated and needed a new intubation within 48 hours (EF, Extubation Failure,'0') and patients who were successfully extubated and did not need reintubation within 48 hours (ES, Extubation Success,'1').

2.2. Data preprocessing

Data cleaning was performed to manage missing data and potential outliers. First, variables or patients with more than 70% of missing values were excluded as well as variables with many outliers. Then, for the remaining variables, we performed imputation of the missing values based on the kNN (k Nearest Neighbors) method, considering the patients of the two classes separately.

A further preprocessing step was applied to reduce the number of variables associated to the patient comorbidities. Starting from the initial 29 binary variables related to the comorbidities, we removed those variables equal to "0" for all patients included in our dataset. Then, we transformed the remaining binary variables into a decimal variable.

After data cleaning and preprocessing, we obtained a dataset containing 13688 patients and 58 variables. The dataset was highly imbalanced: the ES class contained 10832 patients whereas the other 2856 belonged to the EF class.

3. Feature Selection

3.1. Training set construction

The obtained dataset was divided into three sets. First, we randomly extracted 100 ES patients and 50 EF patients to be included in the Test Set (TS). For both groups, 50% of the patients were without imputed values.

We decided to use 2300 patients (1150 for each class) to construct a balanced Training Set (TRS). To this purpose, the remaining 13538 patients were divided according to the class and clustered using agglomerative hierarchical clustering. Then, we proportionally extracted, from the obtained clusters and for each class, 1000 patients containing imputed values and 150 without missing values. A Validation Set (VS) made of 11238 patients remained available to evaluate the feature selection results.

3.2. GA implementation

We chose the Deep Neural Network (DNN) as classification method to identify the weaning outcome. Dimensionality reduction was performed by means of feature selection to preserve the features meaning for explainability purposes. To optimize the subset of features to be used as input of the DNN, we decided to use a Genetics Algorithm (GA).

Each GA solution was made of 58 yes/no (binary) variables corresponding to the features, where 1 means the features is selected, 0 not selected.

The initial GA population consisted of 5000 solutions. At each iteration, 2000 parents were selected using the roulette wheel selection method. To each couple of parents was applied a crossover with 4 cutting points and probability equal to 1, followed by mutation with probability equal to 0.4.

The GA fitness function was based on the DNN classification results. After testing several DNN configurations with different numbers of hidden layers, we defined a triangle DNN structure in which the input layer consisted of a number of neurons equal to the number of selected features, the output layer was made of 1 neuron and the numbers of neurons of each hidden layers was progressively decreased by 4 with respect the previous layer.

The fitness function, to be minimized, was computed using the following formula:

$$fitness = 1 - \frac{TP+TN}{TP+TN+FP+FN} + 0.3 * \left| \frac{TP}{TP+FN} - \frac{TN}{TN+FP} \right| \quad (1)$$

where TP, TN, FP and FN are respectively true positives, true negatives, false positives and false negatives. This fitness equation aimed to prefer solutions with high accuracy but, at the same time, to avoid too much imbalance between sensitivity and specificity.

For each solution, we trained a DNN using the TRS and selecting only the subset of features identified in the current solution. Then, we used the trained network to classify the VS elements and calculate the corresponding fitness value. If the percentage of patients correctly classified was lower than 10%, the current network was discarded and the training was repeated up to a maximum of 5 times. If all the trained DNNs returned unsatisfactory results, the fitness value was set to 1000 to be sure to discard the actual solution in the next generations.

4. Results

4.1. Data cleaning results

The 91 variables were checked to find missing values. From this analysis it emerged that 35 variables had no missing values, 34 variables had less than 1% of missing values, 13 variables had a number of missing between 1% and 40%, 5 variables had missing values between 40% and 70%, and only 4 variables had more than 70% of missing values. The number of missing was decreased removing 3 patients that presented a huge number of variables with no value associated. Moreover, we decided to remove the 4 variables with more than 70%. We also removed height and weight with 32% and 17% of missing values respectively, because the imputation of these variables could be affected by a high

uncertainty and could introduce biases in the classifier construction. After data cleaning, the number of remaining variables was 85. The number of outliers was negligible.

4.2. GA results

Due to the stochastic nature of the algorithm, we decided to perform 50 repetitions of GA starting from the same initial population. For each repetition, we saved the solution with the best (lowest) fitness value. Analyzing the entire set of 50 best solutions, the number of select features ranged from 21 to 39, with a median value of 30. Figure 1 shows, for each variable, how many times it was selected. Four variables were included in most of the solutions (≥40 times): 'Sedation_days', 'Mean_Airway_Pressure', 'PaO2', and 'Chloride'. All variables were selected at least 8 times in the 50 solutions, meaning that all of them are somehow informative in combination with others for the weaning outcome prediction.

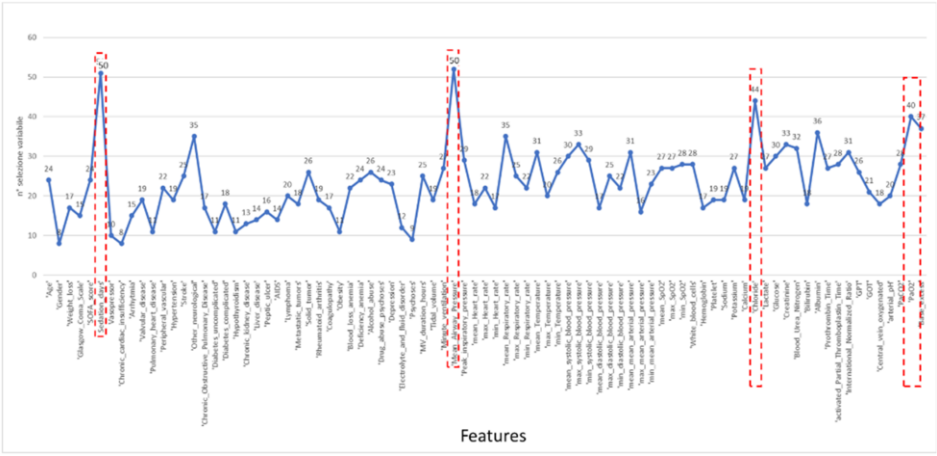


Figure 1. This diagram reports the number of solutions containing each variable.

To assess the goodness of the 50 GA solutions, we analyzed the classification results obtained using the corresponding DNN. We did not perform any tuning on the DNN implementation. Figure 2 shows sensitivity and specificity obtained, on the training and validation sets, using each best solution. Specificity is, on the average, slightly higher than sensitivity considering the TRS (0.72±0.05 vs 0.68±0.06, respectively), whereas on the VS the results are similar (0.69±0.05 vs 0.68±0.06 for specificity and sensitivity respectively). This behavior is essentially due to the penalty term inserted in the fitness function that allows to prefer solutions with classification performance balanced between the two classes. Moreover, it can be observed that the performances on the TRS and VS are quite similar, highlighting the generalization capability of the classifier.

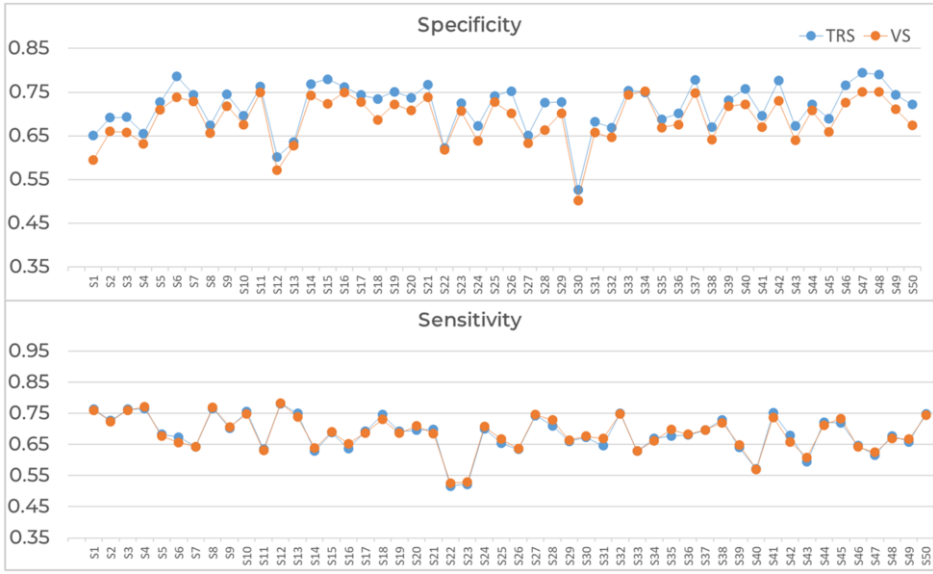


Figure 2. Sensitivity and specificity associated to each GA best solution.

5. Conclusions

In this work we focalized on the role of different kinds of features to predict the outcome of weaning from a mechanical ventilator. Looking at how many times each variable occurs in the best GA solutions, we can state that all features are somehow informative, but four of them are almost essential for this purpose. It is also important to notice that there are different groups of variables associated with them giving similar results. This is only an initial step to obtain a tool to be added to the other clinical indices for minimize the risk of extubation failure.

References

[1] Karthika M, Al Enezi FA, Pillai L V, Arabi YM. Rapid shallow breathing index. Ann Thorac Med 2016;11:167–76. https://doi.org/10.4103/1817-1737.176876.

[2] Fabregat A, Magret M, Ferré JA, Vernet A, Guasch N, Rodríguez A, et al. A Machine Learning decision-making tool for extubation in Intensive Care Unit patients. Comput Methods Programs Biomed 2021;200:105869. https://doi.org/https://doi.org/10.1016/j.cmpb.2020.105869.

[3] Getzen E, Ungar L, Mowery D, Jiang X, Long Q. Mining for Equitable Health: Assessing the Impact of Missing Data in Electronic Health Records. J Biomed Inform 2023:104269. https://doi.org/https://doi.org/10.1016/j.jbi.2022.104269.

[4] Tsai T-L, Huang M-H, Lee C-Y, Lai W-W. Data Science for Extubation Prediction and Value of Information in Surgical Intensive Care Unit. J Clin Med 2019;8. https://doi.org/10.3390/jcm8101709.

[5] Zeng Z, Tang X, Liu Y, He Z, Gong X. Interpretable recurrent neural network models for dynamic prediction of the extubation failure risk in patients with invasive mechanical ventilation in the intensive care unit. BioData Min 2022;15:21. https://doi.org/10.1186/s13040-022-00309-7.

[6] Johnson A, Pollard T, Mark R. MIMIC-III Clinical Database (version 1.4). PhysioNet. 2016. Available from: https://doi.org/10.13026/C2XW26.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230207

Towards an Explainable AI-Based Tool to Predict Preterm Birth

Ilias KYPARISSIDIS KOKKINIDIS[a], Evangelos LOGARAS[a], Emmanouil S. RIGAS[a,1], Ioannis TSAKIRIDIS[b], Themistoklis DAGKLIS[b], Antonis BILLIS[a,2] and Panagiotis D. BAMIDIS[a,2]
[a] *Lab of Medical Physics and Digital Innovation, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Greece*
[b] *3rd Department of Obstetrics and Gynecology, Aristotle University of Thessaloniki, Greece*
ORCiD ID: Ilias Kyparissidis Kokkinidis https://orcid.org/0000-0002-3307-4351
Evangelos Logaras https://orcid.org/0000-0003-4921-3819
Emmanouil S. Rigas https://orcid.org/0000-0002-8042-9135
Ioannis Tsakiridis https://orcid.org/0000-0003-4337-7871
Themistoklis Dagklis https://orcid.org/0000-0002-2863-5839
Antonis Billis https://orcid.org/0000-0002-1854-7560
Panagiotis D. Bamidis https://orcid.org/0000-0002-9936-5805

Abstract. Preterm birth (PTB) is defined as delivery occurring before 37 weeks of gestation. In this paper, Artificial Intelligence (AI)-based predictive models are adapted to accurately estimate the probability of PTB. In doing so, pregnant women' objective results and variables extracted from the screening procedure in combination with demographics, medical history, social history, and other medical data are used. A dataset consisting of 375 pregnant women is used and a number of alternative Machine Learning (ML) algorithms are applied to predict PTB. The ensemble voting model produced the best results across all performance metrics with an area under the curve (ROC-AUC) of approximately 0.84 and a precision–recall curve (PR-AUC) of approximately 0.73. An attempt to provide clinicians with an explanation of the prediction is performed to increase trustworthiness.

Keywords. Artificial Intelligence (AI)**,** Machine Learning (ML), Explainable Artificial Intelligence (XAI), Decision Support Systems, Clinical, Premature Birth, Obstetrics and Gynecology (specialty)

1. Introduction

Preterm birth (PTB) is defined as delivery before 37 weeks of pregnancy. One third of PTBs are medically indicated, primarily preeclampsia and/or fetal growth restriction (FGR), and the other two thirds are spontaneous [1]. According to data from 107 countries in 2014, 14.84 million births—or 10.6% of all births—were PTB [2].

PTB increases the risk of both short- and long-term health effects to the neonate and its later life. Regarding short-term effects, neonatal and childhood mortality are both mostly attributed to PTB [3]. Long-term risks for hypertension and other cardiovascular

[1] Corresponding Author: Emmanouil S. Rigas, Medical Physics and Digital Innovation Lab, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Greece. E-mail: erigas@auth.gr.
[2] A. Billis & PD Bamidis share senior authorship

diseases, type 2 diabetes, kidney disease, respiratory difficulties, and developmental disabilities are elevated [4]. PTB is also linked to higher healthcare expenses [5].

Based on the above, novel approaches to reduce the risk of PTB is one of the main goals in obstetric care. The development of new technologically advanced digital tools and interventions can contribute to achieving this goal. The QUiPP application is one such example [6]. Taking into account the pregnant woman's medical history, current pregnancy information, and predictive clinical tests, the QUiPP application is able to predict her likelihood of giving PTB within clinically significant timeframes. In another study [7], authors proposed the PredictPTB model, a deep learning model, similarly able to predict PTB using routinely collected data from electronic health records (EHRs). Moreover, an intelligent mechanism based on the SVM algorithm has been introduced [8], capable of predicting PTB among others.

In this paper, the aim is to predict PTB in an interpretable way using ML algorithms, find the best-performing ones for PTB, then combine them with ensemble methods and a Multilayer Perceptron (MLP) neural network to increase prediction performance. Furthermore, an emphasis is being placed on the interpretability of the models. To provide explainability, Shapley Additive Explanations (SHAP) [9] were used. This is done to increase the trust of the models from clinicians, who are unlikely to trust the diagnostic recommendations of a black box system [10].

2. Methods

A dataset of 375 pregnancies (of which 128 were identified as preterm) was used in this study. The dataset contained 32 features, including demographics, social and medical history, and obstetrics variables. The information was collected pseudonymously by Hippokration General Hospital in Thessaloniki, as part of an ongoing prospective cohort study that was approved by AUTH's Research Ethics and Conduct Committee (94521/2022), in the context of the HosmartAI project.[2] The dataset was collected by four medical professionals over a four-month period (May-August 2022). The dataset underwent preprocessing, including categorizing input features as numerical, ordinal, or nominal, and using one-hot encoding and a label encoder to map certain features into binary vectors or integer values. Some features with high levels of missing values were dropped, and others were imputed with the most frequent or median value. The dataset was also balanced using random under-sampling and Synthetic Minority Oversampling Technique (SMOTE). The final training set used, after the preprocessing and sampling methods were applied contained 120 entries identified as pre-term and 150 entries that were not, whereas in the test set 30 cases were identified as preterm and 45 as not.

To provide the prediction, the following models were tested: Logistic Regression, SVM, Random Forest, XGBoost, and an MLP perceptron from the scikit learn implementation. Various combinations of Voting and Stacking ensembles [11] were also tested to increase performance. Hyperparameters were determined via Bayesian optimization and error analysis. Respectively, regarding the interpretations, SHAP explanations were used. SHAP was chosen because model independence is a feature of SHAP values, which can be used to explain both generally for each model and locally for each prediction in a consistent way in all models.

[2] https://www.hosmartai.eu/

3. Results

3.1. Model evaluation

Six standard evaluation metrics are used to assess the predictions of all the classifiers. These include ROC-AUC, PR-AUC, Recall, Precision, Accuracy, and F1 score. ROC-AUC reflects the best balance between Sensitivity and Specificity whereas PR-AUC the balance between precision and recall. The developed models were able to predict PTB using the predefined variables from the dataset with accuracies of ~69% (SVM), ~57,3% (Logistic), ~73% (XGBoost), ~70% (Random Forest), 78% (Stacking) and ~81% (Voting) (Table 1). This indicates fair discriminative ability in Random Forest and XGBoost, whereas SVM and Logistic Regression were discarded because of low performance, especially at the Recall (SVM ~0.47, Logistic Regression ~0.5) measure. On all models, 5-fold cross-validation was performed to avoid overfitting.

Table 1. Performance metrics of the ML-based prediction models

ML Algorithm	AUC	Accuracy	Recall	Precision	F1 Score
Random Forest	0.83	0.7	0.47	0.78	0.59
XGBoost	0.83	0.73	0.82	0.65	0.65
Voting ensemble	**0.84**	**0.81**	**0.68**	**0.73**	**0.70**
Stacking	0.83	0.78	0.70	0.68	0.69

A graphic representation of the ROC-AUC and PR-AUC of all models can be seen in Figure 1.

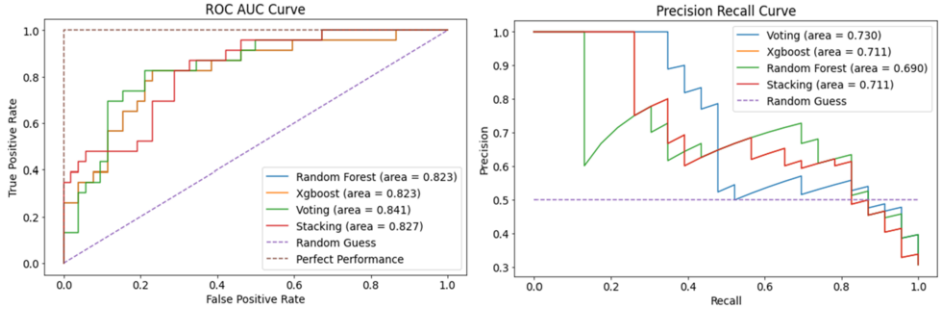


Figure 1. ROC-AUC and PR-AUC for all Classifiers Multi-part figure

In Figure 2 the Confusion Matrix regarding the Voting ensemble Classification algorithm for the preterm labor prediction is displayed. The percentage of the true negative predicted cases is ~57%, and the percentage of the true positive cases is ~22%.

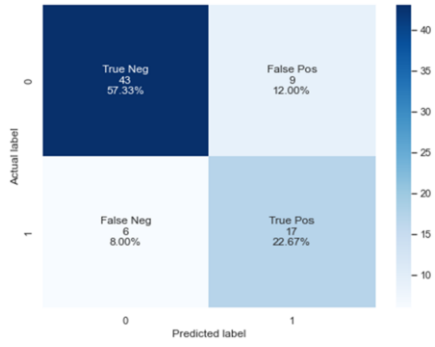


Figure 2. Confusion Matrix for the Voting Classifier

3.2. Feature ranking and Local explanations

In Figure 3(a) we can see the global feature ranking for the predictions of the model. After considering the global explanation for the predicted outcome of PTB in pregnancies, it is also essential to comprehend the output of the models for each specific case. Figure 3(b) illustrates an example of the SHAP explanation for one instance randomly selected from the preterm dataset. This output instance was predicted and confirmed as preterm. Here Placental cord insertion, Pappa, and Gravida play the most important role in the model's output for this instance. It is worth noting that local and global explanations can differ as presented in this instance.

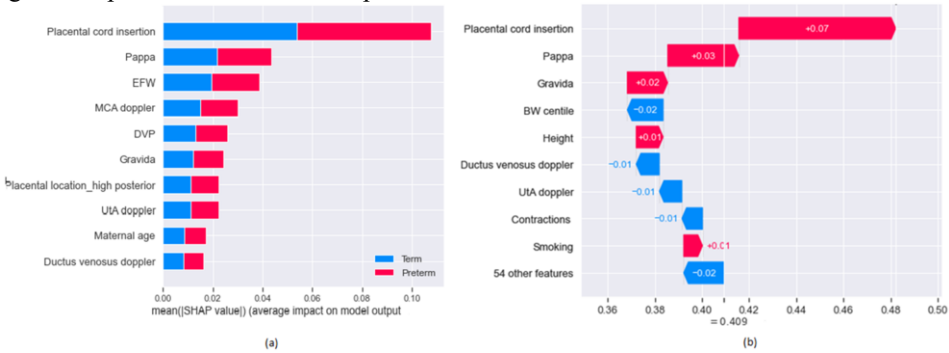


Figure 3. (a) [Global] depicts the influence of features on the constructed classifiers. It illustrates both negative and positive impacts on predictions (red and blue bars). (b)[Local] depicts feature influence on one randomly selected prediction.

4. Discussion and Future work

This study aims to improve machine learning algorithm predictions by providing local explanations that are easy for obstetricians to understand. This can help them make better decisions and provide feedback to improve the model's accuracy in future iterations. The research can also aid in efficient pregnancy screening by giving doctors options and allowing them to understand how each model reaches its decisions.

Given the relatively small dataset that was available for this study, we argue that the initial results that were presented are promising toward the goal of predicting PTB with high accuracy. We observe that currently the Stacking and the Voting ensemble seem to perform better. This is not surprising since ensemble models have been shown to outperform others in diagnostic tasks [12,13]. Moreover, ensemble techniques have been shown to deal better with imbalanced datasets like ours [14]. Compared to previously presented methods for the prediction of PTB, the results in terms of ROC-AUC have been similar. However, we notice an increase of the PR-AUC to ~0.72 compared to other applications [7,8] with the ensemble methods in our models.

Future work will optimize models on a larger dataset and apply 10-fold cross-validation instead of the 5-fold that was employed here to all models.

Acknowledgement

This project has been partially supported by the HosmartAI project – "Hospital Smart development based on AI", which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016834

References

[1] Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet. 2008;371(9606):75–84.
[2] Chawanpaiboon S, et. al., regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. Lancet Glob Health. 2019 Jan;7(1):e37-e46. doi: 10.1016/S2214-109X(18)30451-0. Epub 2018 Oct 30. PMID: 30389451; PMCID: PMC6293055.
[3] Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, Mathers C. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The lancet. 2012 Jun 9;379(9832):2151-61. doi: 10.1016/S0140-6736(12)60560-1.
[4] Raju TN, Pemberton VL, Saigal S, Blaisdell CJ, Moxey-Mims M, Buist S. Long-term healthcare outcomes of preterm birth: an executive summary of a conference sponsored by the National Institutes of Health. The Journal of pediatrics. 2017 Feb 1;181:309-18. doi: 10.1016/j.jpeds.2016.10.015.
[5] Jacob J, Lehne M, Mischker A, Klinger N, Zickermann C, Walker J. Cost effects of preterm birth: a comparison of health care costs associated with early preterm, late preterm, and full-term birth in the first 3 years after birth. The European Journal of Health Economics. 2017 Nov;18(8):1041-6. doi: 10.1007/s10198-016-0850-x.
[6] Dehaene I, Steen J, Vandewiele G, Roelens K, Decruyenaere J. The web-based application "QUiPP v. 2" for the prediction of preterm birth in symptomatic women is not yet ready for worldwide clinical use: ten reflections on development, validation and use. Archives of Gynecology and Obstetrics. 2022 Aug;306(2):571-5.
[7] AlSaad R, Malluhi Q, Boughorbel S. PredictPTB: an interpretable preterm birth prediction model using attention-based recurrent neural networks. BioData Mining. 2022 Feb 14;15(1):6.
[8] Moreira MW, Rodrigues JJ, Marcondes GA, Neto AJ, Kumar N, Diez ID. A preterm birth risk prediction system for mobile health applications based on the support vector machine algorithm. In2018 IEEE International Conference on Communications (ICC) 2018 May 20 (pp. 1-5). IEEE. doi: 10.1109/ICC.2018.8422616.
[9] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30. Retrieved from https://arxiv.org/abs/1705.07874
[10] Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Computer Methods and Programs in Biomedicine. 2022 Sep 27:107161.
[11] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011 Nov 1;12:2825-30.
[12] Sahin H, Subasi A. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. Applied Soft Computing. 2015 Aug 1;33:231-8.
[13] Bashir S, Qamar U, Khan FH. BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. Australasian physical & engineering sciences in medicine. 2015 Jun;38:305-23.
[14] Firdous N, Bhardwaj S. Handling of derived imbalanced dataset using XGBoost for identification of pulmonary embolism—a non-cardiac cause of cardiac arrest. Medical & Biological Engineering & Computing. 2022 Feb 1:1-8. https://doi.org/10.1007/s11517-021-02455-2

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230208

Convolutional Neural Networks for Optical Discrimination Between Histological Types of Colorectal Polyps Based on White Light Endoscopic Images

Vasileios PANTERIS[a,g], Georgios FERETZAKIS[b,c,1], Panagiotis KARANTANOS[a], Dimitris KALLES[b], Vassilios V. VERYKIOS[b], Maria PANOUTSAKOU[a], Eirini KARAGIANNI[a], Christina ZOUBOULI[d], Stefani VGENOPOULOU[d], Aikaterini PIERRAKOU[d], Maria THEODORAKOPOULOU[d], Apostolos E. PAPALOIS[e,g], Thomas THOMAIDIS[f,g], Ilias DALAINAS[h] and Elias KOUROUMALIS[i]

[a] *Gastroenterology Department, Sismanogleio General Hospital, Marousi, Greece*
[b] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[c] *Department of Quality Control, Research and Continuing Education, Sismanogleio General Hospital, Marousi, Greece*
[d] *Pathology Department, Sismanogleio General Hospital, Marousi, Greece*
[e] *Special Unit for Biomedical Research and Education, School of Medicine, Aristotle University of Thessaloniki, Greece*
[f] *2nd Gastroenterology Department, Hygeia Hospital, Athens, Greece*
[g] *Hellenic Society of Gastrointestinal Oncology*
[h] *Administration, Sismanogleio General Hospital, Marousi, Greece*
[i] *Department of Gastroenterology, University of Crete Medical School, Heraklion, Greece*

Abstract. The objective of this study was to compare different convolutional neural networks (CNNs), as employed in a Python-produced deep learning process, used on white light images of colorectal polyps acquired during the process of a colonoscopy, in order to estimate the accuracy of the optical recognition of particular histologic types of polyps. The TensorFlow framework was used for Inception V3, ResNet50, DenseNet121, and NasNetLarge, which were trained with 924 images, drawn from 86 patients.

Keywords. Deep learning, convolutional neural networks, polyps, endoscopy

[1] Corresponding Author: Georgios FERETZAKIS, PhD, School of Science and Technology, Hellenic Open University, Patras, Greece, E-mail: georgios.feretzakis@ac.eap.gr.

1. Introduction and Background

Artificial Intelligence (AI) has already been tested in every aspect of the medical field, namely detection and diagnosis, therapy, and prognosis of diseases. Since the global summit of AI in Gastroenterology held in the USA in 2019, it has been well acknowledged that AI and especially what is called deep learning, will play an important role in the treatment of colorectal polyps [1]. Especially interesting is its application to the histological characterization of these polyps by means of an optical biopsy. This kind of histological determination through the imaging characteristics of a particular polyp has been a conundrum for decades since it has been recognized as a useful asset during colonoscopy either as a screening tool or as a diagnostic modality, for a variety of indications.

Optical biopsy, circumventing the histopathological analysis, is considered important during colonoscopy because it permits the immediate and direct decision of the endoscopist on the risk level of the lesion and consequently on the need for excision or selection of the best endoscopic method for removal, following a cost-effective approach. This approach aims to decrease complications and at the same time to save money, time, and effort for any healthcare system that is based on restricted human and economic resources [2].

Until now the goal of a reliable optical biopsy was concentrated on the technological achievements of better image quality and analysis that will offer better discrimination of the anatomical structures of the polyps during real-time endoscopy, either by means of magnification endoscopes or virtual chromoendoscopy methods [e.g NBI (narrow band imaging)]. These approaches demand specialized and expensive equipment and time-consuming training of experts, which will always depend on the experience and skills of the operator, rendering the whole process mainly subjective, although several classifications have been perpetually adapted in order to reduce bias [3,4]. As a consequence of the trade-off between expectations and necessity, it was considered that artificial intelligence could provide a second opinion to the endoscopist during routine clinical practice for the optical determination of colorectal polyps, probably without the need for the additive cost of endoscopic accessories, while at the same time being feasible even for novice endoscopists who lack the experience of a trained eye [5,6]

The scope of this study was to compare different convolutional neural networks (CNN), as employed in deep learning (DL) algorithms, used on white light images of colorectal polyps acquired during the process of a colonoscopy, in order to estimate the accuracy of the optical recognition of particular histologic types of polyps. Large bowel carcinogenesis has been proven to be a sequential process starting from low-risk lesions progressively transformed to higher risk and finally to cancer. The purpose of a colonoscopy is to prevent this escalation by detecting and removing precancerous polyps to reduce the incidence and mortality of colon cancer. This approach has been proven effective when applied to screening programs that have been addressed to millions of eligible persons [7]. On the other hand, the workload and interventional endoscopies that are produced are so immense that a tool, which will have the ability to accurately classify these polyps, can be expected to reduce the burden on clinical, laboratory, and human resources. This study has also included in the analysis, besides the common precancerous lesions called adenomas, the also important but less investigated and more difficult to characterize sessile serrated adenomas (SSA/P), which constitute the alternative carcinogenesis pathway [8].

2. Methods and Materials

This study was conducted in a public tertiary care hospital in Greece and has been approved by the corresponding Institutional Review Board. The patients who were admitted to the endoscopy unit of a tertiary hospital for the execution of a colonoscopy for several indications were prospectively included in the study except for chronic idiopathic bowel disease and syndromes of polyposis. The individuals were submitted to colonoscopy following routine clinical practice and during the procedure, any polyp that was discovered was photographed using high-definition imaging under white light endoscopy (WLE) with as much clarity as possible avoiding bubbles, feces, blur vision, etc., and stored with a resolution of 720X576 pixels. Then, the polyps were fully resected, and every specimen was put in a different vial for the purpose of cross-referencing between the histopathological report and the specific lesions. This report was used as the gold standard for the characterization of any polyp. The study includes the training process of the CNNs which were provided with polyp images under two main categorization schemes. The first one was a 2-class division between neoplastic (adenomas with low and high-grade dysplasia) and non-neoplastic (hyperplastic) polyps and the second was a 3-class division between adenomas vs hyperplastic vs sessile serrated adenomas (SSA/P).

The Tensorflow framework was used for Inception V3, ResNet50, DenseNet121, and NasNetLarge, which were trained with 924 collected images, drawn from 86 patients, after their pre-processing based on the ImageNet database through a transfer learning process. During training, the same hyper-parameters were set and used for all the networks, including learning rate, batch, epochs, etc. The performance of the CNNs on the histological classification process was evaluated based on Accuracy, Loss as well as the areas under the curves (AUC/ROC curves). We present the results of the best two CNNs that provide the highest accuracy with the lesser loss and the best probability of the right selection during their prediction according to ROC curves.

3. Results

There were 86 patients with a median age of 64 years (range 24-92) and 55% of them were male. 53% of the patients were submitted to colonoscopy to search for polyps either as screening, surveillance, or an indication for polyp excision. The patients carried 191 polyps in total with a median size of 8mm (1st-3rd interquartile: 6-10mm). These polyps were 55% adenomas, 22% hyperplastic, and 17% sessile serrated lesions. There were 924 images captured and stored for analysis without any further enhancement.

The CNNs with the most accurate prognosis were the DenseNet121 and Inception v3 in both the categorization schemes (Table 1). The higher accuracy in both categories was achieved by DenseNet 121 and estimated at 84.7% (Figure 1) and 73.6% respectively.

Table 1. Evaluation metrics of the two best-performing CNNs for the optical classification of colorectal polyps in two categories. 1st category: neoplastic vs non-neoplastic, 2nd category: adenomas vs hyperplastic vs SSA/P

Metric/Method	1st category		2nd category	
	DenseNet121	Inception v3	DenseNet121	Inception v3
Accuracy	84.7%	77.8%	73.6%	70.8%
Loss	0.631	0.848	0.999	1.193
AUC ROC	0.886	0.826	0.815	0.852

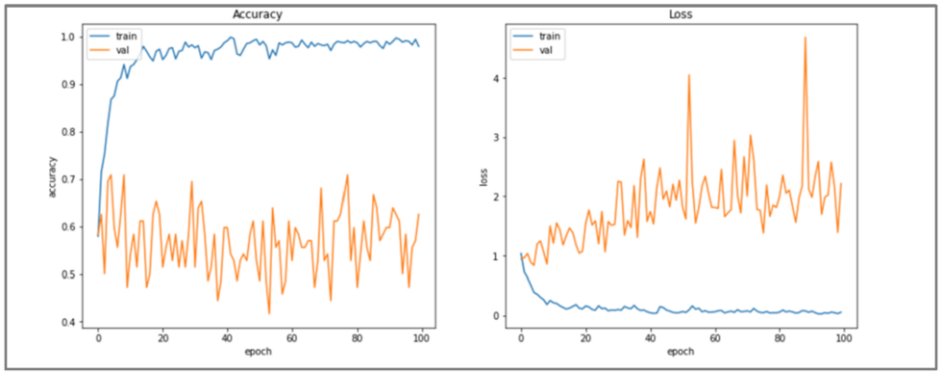


Figure 1. The epoch versus accuracy and loss plots of the DenseNet 121 model on 1st category on train and validation datasets.

4. Discussion

The ability of CNNs to provide a classification for an optical biopsy using white-light endoscopy images of colorectal polyps has been presented in several retrospective studies rendering an accuracy between 67.3-94.4% [3, 9-11]. This wide range of efficacy is probably attributed to the type of CNNs trained and the modulation of their hyper-parameters by each of the different research groups. Furthermore, the actual number of images and the methods to augment their variety and quantity by means of image enhancement as well as the preselection of CNNs with transfer learning might have made a considerable impact on the outcome. Additionally, the type and classes of histologic categorization of polyps, which are utilized by each research group to train the model are crucial, and probably affect the level of difficulty in discriminating certain histological types with respect to others, based on their image characteristics.

In our study, the images have been prospectively collected restricting any selection bias and apart from the classical categorization between neoplastic and non-neoplastic polyps, a venture has been undertaken to include in one of its classifications the SSA/P adenoma for which there have not been enough data in WLE so far. The latter lesions are notorious for their high detection missing rate during colonoscopy, which is attributed to their morphology, location, structure, and size. Furthermore, these lesions have presented dilemmas in their histopathological recognition as the interobserver variability and constantly changing nomenclature have recently indicated. This study has proven a satisfactory accuracy of 84.7% for the DenseNet121 upon WLE during routine clinical practice conditions to select neoplastic lesions and still maintaining a very promising accuracy of 73.6% for discriminating between adenomas, SSA/P, and non-neoplastic lesions. This model seems quite robust demonstrating a quite acceptable level of error gradient and a quite confident AUC of 0.886. This level of accuracy is also comparable to that of expert endoscopists who use image-enhanced endoscopy (IEE) like NBI (accuracy 85.6%) or BLI (accuracy 75%), as prospective and multicenter studies have shown [12,13]. We stress that our results have been achieved without any image enhancement or advanced endoscopic systems.

Admittedly, our study has some limitations. The results are preliminary in the sense that this is an ongoing trial, which is still at the beginning of training the model. The

number of images is still growing, and we expect to enrich our database for further validating the models. Although the training of the model is based on WLE images which can be considered inferior in quality and analysis of specific features, it was the first priority to reflect the capabilities of CNNs under real-world endoscopic conditions that exist in most endoscopy units.

5. Conclusion

The utilization of convolutional neural networks could become useful in the future as an adjunct tool during real-time decision-making of optical histological determination of colorectal polyps by the endoscopists since they could provide acceptable accuracy that at least reach human experts and, at the same time, present promising outcomes even for the most difficult endoscopic lesions. If these results are confirmed by the progress of this study or other research, then artificial intelligence would be able to be applied in everyday clinical practice, thus reducing the impact of disease and the human effort to address it.

References

[1] Kröner PT, Engels MM, Glicksberg BS, Johnson KW, Mzaik O, van Hooft JE, et al. Artificial intelligence in gastroenterology: A state-of-the-art review. World J Gastroenterol. 2021 Oct28;27(40):6794-6824.
[2] Zorron Cheng Tao Pu L, Maicas G, Tian Y, Yamamura T, Nakamura M, Suzuki H, et al. Computer-aided diagnosis for characterization of colorectal lesions: comprehensive software that includes differentiation of serrated lesions. Gastrointest Endosc.2020 Oct;92(4):891-899
[3] Choi SJ, Kim ES, Choi K. Prediction of the histology of colorectal neoplasm in white light colonoscopic images using deep learning algorithms. Sci Rep. 2021 Mar 5;11(1):5311
[4] Kim KO, Kim EY. Application of Artificial Intelligence in the Detection and Characterization of Colorectal Neoplasm. Gut Liver. 2021 May 15;15(3):346-353
[5] Taghiakbari M, Mori Y, von Renteln D. Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. World J Gastroenterol. 2021 Dec 21;27(47):8103-8122
[6] Parsa N, Byrne MF. Artificial intelligence for identification and characterization of colonic polyps. Ther Adv Gastrointest Endosc. 2021 Jun 29;14:26317745211014698. doi: 10.1177/26317745211014698
[7] Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. N Engl J 2012; 366:687–96.
[8] Nishizawa T, Yoshida S, Toyoshima A, Yamada T, Sakaguchi Y, Irako T, et al. Endoscopic diagnosis for colorectal sessile serrated lesions. World J Gastroenterol. 2021 Apr7;27(13): 1321-1329
[9] Yang YJ, Cho BJ, Lee MJ, Kim JH, Lim H, Bang CS, et al. Automated Classification of Colorectal Neoplasms in White-Light Colonoscopy Images via Deep Learning. J Clin Med. 2020 May 24;9(5):1593
[10] Wang Y, Feng Z, Song L, Liu X, Liu S. Multiclassification of Endoscopic Colonoscopy Images Based on Deep Transfer Learning. Comput Math Methods Med. 2021 Jul 3;2021:2485934.
[11] Ozawa T, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therap Adv Gastroenterol. 2020 Mar 20;13:1756284820910659.
[12] Houwen BBSL, Hassan C, Coupé VMH, Greuter MJE, Hazewinkel Y, Vleugels JLA, et al.Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy. 2022 Jan;54(1):88-99
[13] Rees CJ, Rajasekhar PT, Wilson A, Close H, Rutter MD, Saunders BP, East JE, et al. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. Gut. 2017 May;66(5):887-895

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230209

Blood Vessel Segmentation Using U-Net for Glaucoma Diagnosis with Limited Data

Lukas SCHIESSER[a,1], Jens Julian STORP[b], Kemal YILDIRIM[c], Julian VARGHESE[c] and Nicole ETER[b]
[a] *University of Osnabrück, Germany*
[b] *Department of Ophthalmology, University of Münster Medical Center, Germany*
[c] *Institute of Medical Informatics, University of Münster, Münster, Germany*

Abstract. Glaucoma is one of the leading causes of blindness worldwide. Therefore, early detection and diagnosis are key to preserve full vision in patients. As part of the SALUS study, we create a blood vessel segmentation model based on U-Net. We trained U-Net on three different loss functions and used hyperparameter tuning to find their optimal hyperparameters for each loss function. The best models for each of the loss functions achieved an accuracy of over 93%, Dice scores around 83% and Intersection over Union scores over 70%. They each identify large blood vessels reliably and even recognize smaller blood vessels in the retinal fundus images and thus pave the way for improved glaucoma management.

Keywords. Glaucoma, U-Net, Deep Learning, Ophthalmology, Blood vessel segmentation, segmentation

1. Introduction

Glaucoma is one of the most common chronic eye diseases worldwide, causing irreversible visual field defects and blindness. One of the major risk factors for glaucoma progression is an elevated intraocular pressure (IOP) [1]. In medical practice, monitoring of chronic diseases usually requires multiple, repetitive measurements over long time periods in order to adequately evaluate the activity level of a disease and to adjust treatment schemes accordingly. SALUS ("*Selbsttonometrie und Datentransfer bei Glaukompatienten zur Verbesserung der Versorgungssituation*") is a two-arm, multicenter, randomized clinical trial evaluating the state of medical care of glaucoma patients in Germany [2]. In short, glaucoma patients were randomly assigned to one of two groups: 1. inpatient cohort, 2. outpatient cohort. The inpatient group received standard Goldmann applanation tonometry over the course of three consecutive days to assess IOP. Individuals in the outpatient cohort were given iCare Home devices [3], which they used to assess IOP by themselves. The SALUS trial will follow-up on participants until May 2023 and aims to publish results of the comparison of both study arms by the end of 2023. As part of this study, a multi-step deep learning framework is created to assist ophthalmologists with the diagnosis of glaucoma. Although glaucoma is often associated with increased IOP, it is also important to look at physiological changes in the retina. As one-third to half of all patients with early-stage glaucoma

[1] Corresponding Author: Lukas Schießer, E-mail: lschiesser@uos.de.

experience progressing visual field loss despite having adequate IOP compared to control groups [4]. Therefore, non-IOP-dependent methods like evaluating the retinal vasculature are needed since vascular changes can be observed in early stages in the glaucoma's pathogenesis [5]. The segmentation of blood vessels in the papillary region of the retina can aid ophthalmologists in this task. This research aims to contribute to non-IOP-dependent methods by developing models for blood vessel segmentation as means of observing vascular changes more efficiently.

2. Methods and Material

Data. We used the DRIVE dataset [6] to train and test our models. It contains 40 images taken from the retinas of human subjects with corresponding expert generated ground truth segmentation maps indicating the location of blood vessels in each image. It is split into 20 training and 20 test images. The dataset is commonly used in medical image analysis and has been widely used to research automated blood vessel segmentation. We additionally used 5 random images from the STARE dataset [7] as a validation dataset for hyperparameter tuning.

Preprocessing and Data Augmentation. All images were first preprocessed using contrast limited adaptive histogram equalization (CLAHE). Since the dataset is very small, we enlarged the dataset by using different combinations of image augmentation techniques, namely, rotating the images by at most 10° as well as changing the brightness, contrast, and saturation of the fundus images using the albumentations package [8]. Thus, we created 60 additional images for training, totaling 80 training images. No image augmentation techniques were applied to the test and validation images.

U-Net is a powerful convolutional neural network commonly used for biomedical image segmentation [9]. Its architecture is based on an encoder-decoder network. The encoder which is also sometimes called the *contractive path* downsamples the input image, while the decoder, also sometimes called the *expansive path,* upsamples the encoded representation producing a segmentation mask. U-Net's encoder and decoder networks are connected by a series of skip connections which help to preserve spatial information and improve accuracy of the segmentation. These connections pass information from the encoder to the decoder at multiple scales which allows the decoder to make more accurate predictions by taking into account both global and local information from the input image [10]. U-Net can already achieve results comparable with sliding-window based convolutional networks when trained on an extremely small dataset of a few hundred images. When adding data augmentation, preprocessing and enhancement techniques, it outperforms existing state-of-the-art methods on several biomedical image segmentation challenges [9]. A pretrained architecture can be employed as U-Net's encoder path to improve the model's general performance as measured by accuracy and *Intersection over Union* (IoU) score [11]. We chose a ResNet-34 architecture pretrained on the ImageNet dataset as our encoder path.

Training. To segment the blood vessels, we employed a U-Net network with a ResNet-34 encoder pretrained on the ImageNet dataset using the Segmentation Models package [12]. We decided to use a mini-batch approach. We further used the Adam optimizer with a learning rate of 0.001. We compared three different loss functions for training the model, namely, Tversky Loss, Dice Loss, and Focal Loss. To find the best-performing versions of our model for all three loss functions, we employed

hyperparameter tuning using Ray Tune [13] during which we also investigated the effect of the number of epochs on the maximum IoU score of our models.

Hyperparameter tuning. As already mentioned, we decided to use the number of epochs and the alpha values of the Tversky and Focal Loss for hyperparameter tuning. The Dice Loss has no tunable hyperparameters. The Tversky Loss's alpha value is used to weigh the relative importance of false positives and false negatives. Since blood vessel segmentation is a very complex task due to the widely ranging sizes of blood vessels in the retina, we want to penalize false negatives more than false positives [14]. A similar reasoning can be applied to the Focal Loss's alpha value. They weigh the different classes of the segmentation task. Since we were more interested in the correct identification of our foreground class (the blood vessels) than the background class, we weighed them more strongly. For finetuning the number of training epochs, we decided to keep the number of epochs trained relatively narrow to avoid overfitting and keep training times short. An overview of which hyperparameter values were tested for which loss can be found in Table 1. The hyperparameter tuning was evaluated for all three loss functions separately. We searched the hyperparameter space with respect to maximizing IoU scores using grid search evaluating the performance of each model on the validation dataset.

Table 1. All possible hyperparameters used during hyperparameter tuning.

Loss function	Hyperparameter	Range
Tversky loss	Epochs	10, 15, 20, 25, 30, 35, 40
	Alpha	0.5, 0.6, 0.7, 0.8, 0.9
Dice loss	Epochs	10, 15, 20, 25, 30, 35, 40
Focal loss	Epochs	10, 15, 20, 25, 30, 35, 40
	Alpha	0.5, 0.6, 0.7, 0.8, 0.9

3. Results

After hyperparameter tuning, we used the best hyperparameters to retrain the models from scratch. The best Tversky Loss combination was trained for 25 epochs with an alpha value of 0.6. It achieved an IoU score of 0.8674 on the validation data. The best Dice Loss combination was trained for 35 epochs and achieved an IoU score of 0.8639 on the validation data. Finally, the Focal Loss performed best during the tuning process with an IoU score of 0.8715. Its alpha value was 0.5 which implicitly results in an 0.5 for the second class in the weighting vector. The model was trained for 30 epochs. The models were then evaluated on a held-out test dataset. The model trained with the Focal Loss function performed best with an IoU score of 0.718. Table 2 includes all test metrics computed for the three different loss functions with their best hyperparameter combination.

Table 2. Best Hyperparameter combination and corresponding metrics during testing

Model	Hyperparameter Configuration	IoU score	Dice Score	Accuracy
Tversky Loss	Alpha: 0.6, Epochs: 25	0.705	0.826	0.929
Dice Loss	Epochs: 35	0.712	0.830	0.931
Focal Loss	Alpha: 0.5, Epochs: 30	0.718	0.836	0.930

4. Discussion

By using hyperparameter tuning, we created three models with solid performance on the task of blood vessel segmentation in fundus images. All three models achieve an accuracy of around 93%. However, an IoU score of just over 70% and a dice score of just over 82% is not necessarily considered satisfactory. Nevertheless, these results still attest a good performance of our models, especially given our small dataset. Looking at the predicted blood vessel segmentation masks and comparing them to the ground truth masks (see Figure 1), we can visually confirm a good performance.

Nevertheless, the reliability of our validation metrics on which we base our decision during hyperparameter tuning on might be subject to improvements. Since we only used five images during validation, we cannot count on the same statistical properties that we normally count on with larger validation datasets. Therefore, we must regard our validation metrics more carefully than with larger datasets. This could be avoided by creating datasets with a larger number of data points, e.g., by merging multiple datasets together, acquiring more images with ground truth masks.

Another error source could be the vast amount of image augmentation that was used to increase the size of the training dataset. By applying multiple augmentation techniques, we increased our training data size from 20 to 80. This can lead to overfitting and a worse performance [15]. Further experiments with differing amounts of image augmentation will be conducted in follow-up studies.

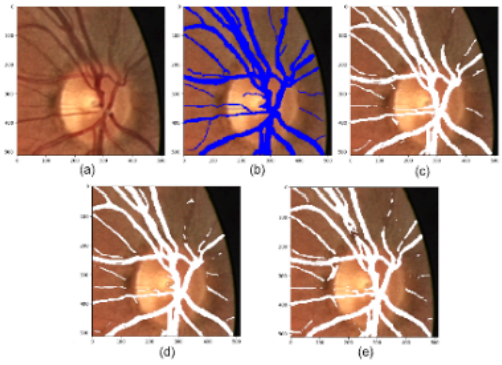


Figure 1. Comparison of predicted segmentation masks (white) with the ground truth mask (blue). (a) Input image to the model (preprocessed with CLAHE) (b) Ground truth mask (c) Segmentation mask predicted by model trained with Dice Loss (d) Segmentation mask predicted by model trained with Tversky Loss (e) Segmentation mask predicted by model trained with Focal Loss

5. Conclusion and Outlook

Our experiments show that U-Net can accurately segment blood vessels, even with small datasets. Our models perform comparably to other approaches and correctly identify major blood vessels in the papillary region while also segmenting smaller vessels reliably. While segmenting the entire retinal fundus image would yield better results, we are constrained by our larger multi-step framework for diagnosing glaucoma as part of the SALUS study.

In the future, we could extend our U-Net model to include data about arterial and venous blood vessels, which might be advantageous during the diagnosis process as

Chan et al. suggest [5]. The RITE dataset provides ground truth masks for the DRIVE dataset's images with these classes.

References

[1] Weinreb RN, Aung T, Medeiros FA. The Pathophysiology and Treatment of Glaucoma: A Review. JAMA [Internet]. 2014 [cited 2022 Nov 30];311:1901. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2014.3192.

[2] Oldiges K, Steinmann M, Duevel JA, et al. SALUS—a non-inferiority trial to compare self-tonometry in glaucoma patients with regular inpatient intraocular pressure controls: study design and set-up. Graefes Arch Clin Exp Ophthalmol [Internet]. 2022 [cited 2022 Nov 30];260:3945–3955. Available from: https://link.springer.com/10.1007/s00417-022-05759-7.

[3] Valero B, Fénolland J-R, Rosenberg R, et al. Fiabilité et reproductibilité des mesures de la pression intraoculaire par le tonomètre Icare® Home (modèle TA022) et comparaison avec les mesures au tonomètre à aplanation de Goldmann chez des patients glaucomateux. Journal Français d'Ophtalmologie [Internet]. 2017 [cited 2022 Nov 30];40:865–875. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0181551217303984.

[4] Gupta N, Weinreb RN. New definitions of glaucoma. Curr Opin Ophthalmol. 1997;8:38–41.

[5] Chan KKW, Tang F, Tham CCY, et al. Retinal vasculature in glaucoma: a review. BMJ Open Ophthalmol [Internet]. 2017 [cited 2022 Dec 5];1:e000032. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5721639/.

[6] Staal JJ, Abramoff MD, Niemeijer M, et al. Ridge based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging. 2004;23:501–509.

[7] Hoover AD, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Transactions on Medical Imaging. 2000;19:203–210.

[8] Buslaev A, Iglovikov VI, Khvedchenya E, et al. Albumentations: Fast and Flexible Image Augmentations. Information [Internet]. 2020;11. Available from: https://www.mdpi.com/2078-2489/11/2/125.

[9] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [Internet]. arXiv; 2015 [cited 2022 Jun 29]. Available from: http://arxiv.org/abs/1505.04597.

[10] Chen L-C, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [Internet]. arXiv; 2018 [cited 2022 Dec 6]. Available from: http://arxiv.org/abs/1802.02611.

[11] Iglovikov V, Shvets A. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation [Internet]. arXiv; 2018 [cited 2022 Dec 7]. Available from: http://arxiv.org/abs/1801.05746.

[12] Iakubovskii P. Segmentation Models Pytorch [Internet]. GitHub repository. GitHub; 2019. Available from: https://github.com/qubvel/segmentation_models.pytorch.

[13] Liaw R, Liang E, Nishihara R, et al. Tune: A Research Platform for Distributed Model Selection and Training. arXiv preprint arXiv:180705118. 2018;

[14] Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks [Internet]. arXiv; 2017 [cited 2022 Dec 10]. Available from: http://arxiv.org/abs/1706.05721.

[15] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data [Internet]. 2019 [cited 2022 Dec 11];6:60. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230210

First Steps Towards a Risk of Bias Corpus of Randomized Controlled Trials

Anjani DHRANGADHARIYA[a,b,1] ,Roger HILFIKER[c], Martin SATTELMAYER[d], Katia GIACOMINO[d], Rahel CALIESCH[d], Simone ELSIG[d], Nona NADERI[e,f] and Henning MÜLLER[a,b]
[a] *Informatics Institute, HES-SO Valais-Wallis, Sierre, Switzerland*
[b] *University of Geneva (UNIGE), Geneva, Switzerland*
[c] *IUFRS, University of Lausanne, Lausanne, Switzerland*
[d] *School of Health Sciences, HES-SO Valais-Wallis, Leukerbad, Switzerland*
[e] *Geneva School of Business Administration, HES-SO Geneva, Switzerland*
[f] *SIB Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland*
ORCiD ID: Anjani Dhrangadhariya https://orcid.org/0000-0003-1691-1338

Abstract. Risk of bias (RoB) assessment of randomized clinical trials (RCTs) is vital to conducting systematic reviews. Manual RoB assessment for hundreds of RCTs is a cognitively demanding, lengthy process and is prone to subjective judgment. Supervised machine learning (ML) can help to accelerate this process but requires a hand-labelled corpus. There are currently no RoB annotation guidelines for randomized clinical trials or annotated corpora. In this pilot project, we test the practicality of directly using the revised Cochrane RoB 2.0 guidelines for developing an RoB annotated corpus using a novel multi-level annotation scheme. We report inter-annotator agreement among four annotators who used Cochrane RoB 2.0 guidelines. The agreement ranges between 0% for some bias classes and 76% for others. Finally, we discuss the shortcomings of this direct translation of annotation guidelines and scheme and suggest approaches to improve them to obtain an RoB annotated corpus suitable for ML.

Keywords. risk of bias, annotation, systematic reviews, corpus, automation

1. Introduction

Systematic reviews (SRs) synthesized from randomized controlled trials (RCTs) are the highest quality evidence in the evidence hierarchy and are used by doctors to make diagnostic and treatment decisions. In theory, an RCT accurately measures the treatment effect on patient outcomes but can be biased in practice due to flawed study design, execution, analysis, or outcome reporting [1]. Biases in RCTs cannot be measured, but risk bias can be assessed. So, the reviewers must rigorously look for possible biases before incorporating them into SRs. Published RCTs are exponentially increasing[2], making manual assessment a protracted process. Machine learning (ML) can help

[1] Corresponding Author: Anjani Dhrangadhariya, Informatics Institute, HES-SO Valais-Wallis, Technopole 3, 3960 Sierre, Switzerland; E-mail: anjani.dhrangadhariya@hevs.ch.

[2] https://pubmed.ncbi.nlm.nih.gov/?term=randomized%20controlled%20trial&filter=pubt.randomizedcontrolledtrial

accelerate this process by directly pointing the reviewers to the parts of the text relevant to identifying RoB, leading to quickly judging the trial quality. Automation will thereby accelerate the process of writing SRs, which is tedious and time-consuming. Both Marshall *et al.* and Millard *et al.* attempted automated RoB assessment, albeit using proprietary, pay-walled data [2,3]. Recently, Wang *et al.* released a hand-labelled RoB corpus for preclinical animal studies, not RCTs [4]. RoB assessment of RCTs is a knowledge-heavy task where even highly trained experts are prone to subjective judgments. Developing such a corpus entails creating a clear-cut annotation scheme and guidelines. As neither exists, we focus on two primary concerns: 1) To test whether the widely used revised Cochrane's RoB 2.0 tool for RCTs (RoB 2.0) could be used as RoB annotation guidelines to develop a corpus that could be used for training ML models. 2) To develop and test an RoB annotation scheme that closely mimics the RoB 2.0 [5,6].

2. Methods

2.1. Formulating annotation scheme

RoB 2.0 tool divides biases into five risk domains which further decompose into several signalling questions (SQ), each corresponding to different parts of the trial design. Each signalling question prompts the reviewer to look for a piece(s) of factual evidence in the RCT and, depending on the amount of evidence found to respond with one of the five response options: "Yes", "Probably yes", "No", "Probably no", or "No information". E.g., to respond to the SQ "Was the allocation sequence random?", the reviewers need to identify whether a proper methodology was used for random participant allocation, and only if a proper methodology is identified the reviewer responds to this question as "Yes", and otherwise "No". We formulated an annotation scheme (see Figure 1 where each SQ is an entity. Each entity has five entity labels corresponding to the five response options to that question. Entities represent the factual evidence from the RCTs, and the entity labels incorporate the reviewer's risk judgment.

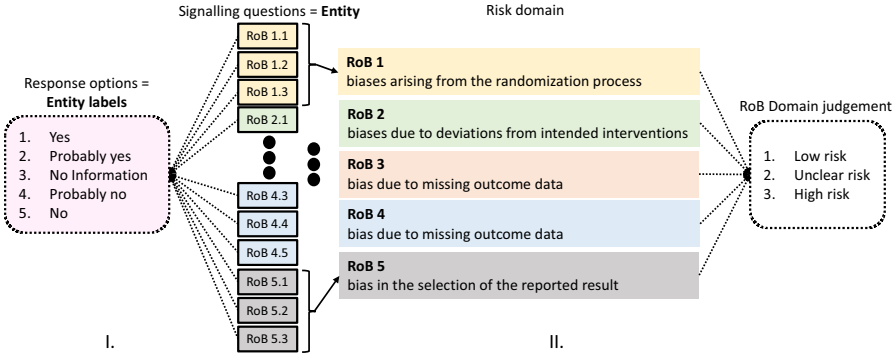


Figure 1. Annotation scheme. I. SQ level: each SQ (RoB 1.1, 1.2, ...) is an entity that could take either of five response options (entity labels). SQ response judgements for individual risk domains (RoB 1-5) could be combined to arrive at risk domain judgement. Note: We do not address risk domain judgments in this work.

2.2. Preliminary annotation guidelines

Full-text RCTs were annotated using the RoB assessment instructions from RoB 2.0.[3] The author, with Natural Language Processing expertise, developed the generic annotation guidelines with four physiotherapists experienced in bias assessment to ensure consistency. Complete sentence(s) or phrase(s) were annotated depending on the text parts relevant to answering an SQ. All the text information pertinent to answering a question was marked, even if the information was found in different parts of the full text. Table or figure captions relevant to answering were marked. If the information was not found in the captions, it was marked within the table contents. If a table or figure reference answered the question, it was annotated.

2.3. Pilot annotation

R.H., M.S., K.G., and R.C. consented to annotate and did the pilot annotation on a corpus of ten RCTs sampled in the following manner. An Entrez[4] search using the search query "`(randomized[title] or randomized[title]) and (rehabilitation or (physical therapy))`" was performed ten times to retrieve studies from one-year timespans, each between 2000 - 2019. Each query was restricted to retrieve 1000 documents, of which ten were randomly chosen for each period. We took the first possible study of the ten sampled studies with a freely available PDF (Portable Document Format). R.H. and M.S. are professors and associate professors, and K.G. and R.C. are doctoral researchers with experience conducting RoB ratings in several SRs. Tagtog[5], a commercial tool, was used for annotating PDFs. The task was to annotate text relevant to answering each signalling question entity and choose a judgment response option entity label. We report the pairwise, token-level F1 that disregards out-of-the-span (unannotated) tokens, which is the ideal measure of annotation reliability for the token-level annotations. [7] F1 is reported for entity IAA_{sq} and entity label $IAA_{response}$ annotations. IAA_{sq} and $IAA_{response}$ measure the reliability of the RoB 2.0 guidelines for selecting the same parts of the text to answer SQs.

3. Results

The pilot annotation resulted in 902 labels corresponding to the SQs and their response options. Table 1 reports pairwise IAA_{sq} and $IAA_{response}$ averaged over all the annotator pairs at the SQ response option level. Individual pairwise IAA_{sq} range between 0% (poor) and 75% (substantial), with most values falling under the poor category and very few under the substantial agreement. SQs RoB 1.1, 1.2, 1.3, 2.6, and 3.1 fared well regarding the average pairwise agreement between all pairs, but none of these categories had a substantial agreement. Questions 2.1, 2.3, 2.4, 2.5, 2.7, 3.4, 4.4, 4.5, and the entire domain 5 fared extremely poorly or with no agreement or annotation. The $IAA_{response}$ scores are considerably lower (to zero) than IAA_{sq}, hinting that annotators choose the

[3] https://drive.google.com/file/d/19R9savfPdCHC8XLz2iiMvL_71lPJERWK/view

[4] The Entrez Global Query Cross-Database Search System is a federated search engine or web portal that allows users to search PubMed database.

[5] https://www.tagtog.com/

same text to answer an SQ but assign different response options to the selected text. The $IAA_{response}$ scores remain variable across the risk domains, with 52.63% of the total scores being zero and no annotation for about 22% of the total scores.

Table 1. Left: Table lists IAA_{sq} between the six annotator pairs (P1-P6)[6] for the RoB SQs. Substantial (≥61) agreements are in bold. Right: Table lists IAA_{sq} averaged over the six annotator pairs for the SQs at the entity label level ($IAA_{response}$). Note Y = Yes, PY = Probably Yes, NI = No Information, N = No and PN = Probably No, Avg. = Average. "-" shows that one of the annotators did not annotate any text for a particular SQ.

SQ	P1	P2	P3	P4	P5	P6	Avg.	Y	PY	NI	PN	N
1.1	23.1	24.5	52.2	57	48	21.5	37.7	21.8	7.1	0	-	-
1.2	**66.1**	50.3	72.8	50.7	46	50.5	56.1	4.9	11.5	10.2	0	-
1.3	**69.5**	20.5	16.1	31.6	59.9	53.5	41.8	-	-	41.8	11.4	9.9
2.1	1	1.4	0	9.1	19.1	0	5.1	8.2	0	-	3	0
2.2	18.3	7.3	11.1	0	23	7.4	11.2	3.6	0	0	0	0
2.3	20.6	5.5	13.4	0	0	0	6.6	-	0	-	1	0
2.4	0	-	-	0	0	-	0	-	0	-	0	-
2.5	0	0	0	0	0	-	0	0	0	-	0	-
2.6	**75.3**	**68.9**	19.3	**63.9**	12.9	19.6	43.3	39.4	0	0	0	3.6
2.7	0	6.6	0	0	0	0	1.1	0	0	-	0	0
3.1	45.8	23.6	32.2	43.4	22.9	14.8	30.4	47.6	0.6	-	1.3	3.3
3.2	1.4	0	0	3.3	7.4	0.9	2.2	0	0	-	0	0
3.3	0	0	0	16.4	0	0	2.8	-	0	31.4	0	0
3.4	-	0	-	0	0	0	0	0	0	0	0	0
4.1	4	6.6	14.2	25.6	22.3	6.3	13.2	-	-	-	0.8	12
4.2	1.8	0	0.4	0	40.1	0	7.1	-	-	-	0.3	0
4.3	7.6	13.9	5	10.5	39.5	8.4	14.2	0	0	0	13.1	20.5
4.4	0	0	0	0	0	0	0	0	0	-	0	0
4.5	0	0	0	0	0	0	0	0	0	-	0	-
5.1	0	0	0	0	0	4.2	0.7	0	0	0	0	0
5.2	23.9	0	0	0	0	2.4	4.4	-	0	0	0	0
5.3	0.2	0	0	0.4	8.1	42	8.4	-	0	0.6	0	0

4. Discussion

We analyzed annotations over all annotator pairs and RoB classes identifying four types of annotation disagreements. A **polarity disagreement** arises when two annotators choose the same chunk of text to answer an SQ but choose polar opposite entity labels ("Yes" or "Probably yes" vs "No" or "Probably no" vs "No information"). In one of the documents, all four annotators chose the same text evidence ("71 allocated routine services, 67 allocated intervention service, ...") to answer SQ 3.1. However, three of the four annotators responded to this question with "Yes", but one chose "Probably no". This SQ asks whether the outcomes data were available for all, or nearly all, participants randomized but does not clarify the exact cut-off for how many participant dropouts increase the risk. Therefore, the annotators make subjective response judgments depending upon what exact percentage of participant dropout is considered valid in their experience. A **degree disagreement** causes low $IAA_{response}$ and arises because some annotators are lenient in judging risk while others are sceptical. The lenient ones select

[6] P1 = R.H. and K.G., P2 = R.H. and M.S., P3 = R.H. and R.C., P4 = K.G. and M.S., P5 = K.G. and R.C., P6 = M.S. and R.C.

definitive "Yes" or "No" for responding to an SQ, while the sceptical ones choose "Probably yes" or "Probably no". A practical and rationally justified solution is to merge the response options "Probably yes" with "Yes" and "Probably no" with "No" to reduce the complexity of the task and increase IAA without altering the final risk judgment for this risk domain. [6] A low IAA is also caused by our annotation guidelines not limiting the annotators to selecting either the phrase vs a sentence(s) vs a paragraph for answering the question leading to a **text span disagreement**. RoB 2.0 tool led to some annotators using and annotating very condensed information to come to a response. In contrast, others used an entire paragraph to reach the same response for an SQ leading to a low token-level IAA. This problem requires mending the annotation guidelines to precisely instruct authors to select the complete information they used to decide or the minimum necessary information to decide on an SQ. Another method is automatically extending the more condensed annotations to the broadest ones. In our guideline improvement, we restrict the annotation to marking the full sentence(s) where the relevant information is found. Sometimes annotators came to a response judgment for an SQ but used different parts of the RCT text leading to **disparate document section disagreement**. Such disagreements emanate because RoB 2.0 do not instruct the annotators about what part of the RCT to annotate and what part to not annotate for a particular SQ. We noticed many SQs remained unanswered because the annotators did not understand what part of the text to annotate, even after following the RoB 2.0 guidelines.

5. Conclusion

In conclusion, the revised Cochrane RoB 2.0 guidelines cannot be directly used as RoB corpus annotation guidelines. It is imperative to develop clear-cut guidelines to instruct the annotators in signalling question and response judgment decisions. The multi-level annotation schema also needs improvement, as discussed. We are using the insights from this pilot annotation to develop detailed, crisp guidelines and obtain consistent annotations. The annotated dataset is available on Zenodo (DOI: 10.5281/zenodo.7698941).

References

[1] Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research. BJOG: an international journal of obstetrics and gynaecology. 2018 Dec; 125(13): p. 1716.
[2] Marshall IJ, Kuiper J, Wallace BC. Automating risk of bias assessment for clinical trials. IEEE journal of biomedical and health informatics. 2015; 19(4): p. 1406--1412.
[3] Millard LA, Flach PA, Higgins JP. Machine learning to assist risk-of-bias assessments in systematic reviews. International journal of epidemiology. 2016; 45(1): p. 266--277.
[4] Wang Q, Liao J, Lapata M, Macleod M. Risk of bias assessment in preclinical literature using natural language processing. Research synthesis methods. 2022; 13(3): p. 368--380.
[5] Lansbury L, Lim B, Baskaran V, Lim WS. Co-infections in people with COVID-19: a systematic review and meta-analysis. Journal of Infection. 2020; 81(2): p. 266--275.
[6] Sterne JA, Savović J, Page M, Elbers R, Blencowe N, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019 Aug; 366.
[7] Deleger L, Li Q, Lingren T, Kaiser M, Molnar K, Stoutenborough L, et al. Building gold standard corpora for medical natural language processing tasks. In AMIA Annual Symposium Proceedings; 2014. p. 144.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230211

Developing a Comprehensive Search Strategy for the Systematic Review of Clinical Decision Support Systems for Nursing Practice

Cynthia ABI KHALIL[a,b,1], Antoine SAAB[c,b], Jihane RAHME[a] and Brigitte SEROUSSI[b,d]
[a]*Nursing Administration, Lebanese Hospital Geitaoui–UMC, Beirut, Lebanon*
[b]*Sorbonne Universite´,Universite´ Sorbonne Paris Nord, INSERM, Laboratoire d'Informatique Me´dicale et d'Inge´nierie des connaissances en e-Sante´, LIMICS, F-75006 Paris, France*
[c]*Quality and Safety Department, Lebanese Hospital Geitaoui–UMC, Beirut, Lebanon*
[d]*AP-HP, Ho^pital Tenon, Paris, France*
ORCiD ID: Cynthia Abi Khalil https://orcid.org/0000-0001-9112-628X
Antoine Saab https://orcid.org/0000-0003-0673-0875
Brigitte Séroussi https://orcid.org/0000-0002-7676-2832

Abstract. The search strategy of a literature review is of utmost importance as it impacts the validity of its findings. In order to build the best query to guide the literature search on clinical decision support systems applied to nursing clinical practice, we developed an iterative process capitalizing on previous systematic reviews published on similar topics. Three reviews were analyzed relatively to their detection performance. Errors in the choice of keywords and terms used in title and abstract (missing MeSH terms, failure to use common terms), may make relevant articles invisible.

Keywords. Nursing informatics, Clinical decision support systems, Systematic reviews, Precision, Recall

1. Introduction

The literature surrounding the clinical use of clinical decision support systems (CDSSs) in healthcare suggests that these systems are effective to some extent. Published studies usually report on CDSSs targeted on physician performance [1], but the number of studies exploring the success of CDSSs in improving nursing clinical practice has increased in the last years [2], showing there is still room for improvement. In order to develop a CDSS with positive impacts on nursing process, we started with a review of the literature. The search strategy of systematic reviews is of utmost importance, as the quality and validity of the review's findings could be directly affected by the completeness and relevance of identified articles. Errors made in the search process may indeed result in a biased or otherwise incomplete list of articles for the review.

[1] Corresponding Author: Cynthia ABI KHALIL, E-mail: cynthiakjammal@outlook.com

While it is almost impossible, although expected, to build a search strategy that identifies all the relevant scientific articles related to a certain scope, researchers thrive towards having the most comprehensive search results. In practice, this is performed through searching various scientific bibliographic databases (BDBs) such as Embase, Cochrane Library, Web of science, MEDLINE, etc. (the latter being the most popular among researchers in the biomedical sciences field) through a query built using a formal syntax specific to each search engine. While a plethora of scientific and educational resources exist regarding methods to correctly query bibliographic databases, the rate of errors in search strategies of reviews is surprisingly high, reaching 90% with at least half of the errors affecting recall [3]. This also affects Cochrane reviews which is considered as the gold standard [4].

The objective of this work is to analyze the search strategy of published reviews related to CDSSs in nursing practice in order to elaborate an optimized comprehensive search strategy for a systematic review on this topic.

2. Methods

First, a Medline/PubMed query was constructed by two of the authors (CAK and BS), in accordance with the PRESS Guideline Evidence-Based Checklist. The question framework and its delineated concepts were defined, and related terms were harvested so as to build a comprehensive list of conceptual synonyms. The query was built by identifying and extracting the controlled vocabulary terms (i.e., MeSH terms) and in-text synonyms (i.e., "common terms") that best expressed the research question and the components of the search segments.

As usual, Boolean operator OR was used between terms relative to the same concept in the search segment, so as to increase sensitivity, and Boolean operator AND was used to combine the search segments in order to improve the specificity of retrieved results. Only articles in English, concerning human research, published between 2000 and 2022, with abstract and full-text available were included. The query initially returned 1581 articles among which we found 23 reviews. Reviews with narrow scope (e.g., use of CDSSs in homecare) or specific to one domain or specialty (e.g., use of CDSSs for pressure ulcers) were excluded to only keep eight reviews [1,2,5–10] with a general scope on CDSSs in nursing. These eight reviews all together had selected a total of 138 articles that we used to revisit the query iteratively, adding and removing terms, to gauge the search strategy to include the articles among the 138 articles that were relevant to our search. We finally reached an *optimized* query that returned a total of 3,283 results.

In a second step, the time frames covered by the eight reviews of interest were analyzed (Figure 1). We finally selected three reviews [2,9,10] for evaluation on a common data extraction period (2014-2017), excluding old reviews (more than 10-year-old). The *optimized* query filtered on the same period (2014-2017) yielded 578 articles. Two of the authors (CAK and AS) independently screened all titles, abstracts, and full texts when needed, to classify articles as *relevant*, *non-relevant* or *may-be relevant* (posters, commentaries, short papers were excluded). Consensus between the two researchers was used to resolve any discrepancies. In total, 145 relevant articles published between 2014 and 2017 were finally included to build the "reference list" (RL).

The performance of each of the three reviews to support our search was assessed through the computation of precision defined as the proportion of retrieved articles

relevant to the review query, and recall defined as the proportion of relevant articles successfully retrieved by the query.

We adopted a proxy measure based on the comparison of articles retrieved by the reviews and articles included in the RL, assumed as the Gold Standard. We considered that articles retrieved by the reviews were relevant when they were also included in the RL (true positives). Thus, we defined precision as the ratio of the number of articles retrieved by the review *and* included in the RL to the number of articles retrieved by the review, and recall as the ratio of the number of articles retrieved by the review *and* included in the RL to the number of articles included in the RL. To get the number of articles included in the RL but missing in the reviews, authors (CAK, AS, and JR) analyzed the full text version of the 145 RL articles taking into account the scope of each review and the inclusion and exclusion criteria that reviews' authors used.

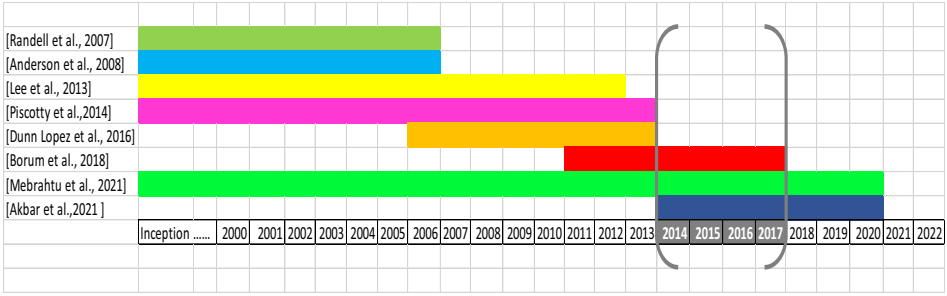


Figure 1. Time frames of reviews to select the largest period that covers the maximum of recent reviews.

3. Results

Akbar et al., [2] retrieved 15 articles in Medline on the [2014;2017] time frame. However, three articles were eliminated because out of the scope of our query (one article had no abstract available, one article was a commentary, and one article was actually a poster) leading to only keep 12 articles for [2]. Thus, the three selected reviews returned 12 [2], 8 [9], and 8 [10] articles in compliance with the inclusion criteria of the optimized query, among which, 11, 5, and 3 articles, respectively, were included in RL.

The optimized query succeeded in retrieving additional 27, 2 and 8 articles published between 2014-2017 and eligible for inclusion for respectively [2], [9] and [10]. Precision on Medline in the selected period is 91.70% (11/12) for [2], 62,50% (5/8) for [9], and 37.50% (3/8) for [10]. In the same way, recall is 28.90% (11/38) for [2], 71.40% (5/7) for [9], and 27.30% (3/11) for [10]. Results are displayed in Table 1.

4. Discussion

Results show that the majority of articles included in reviews are indexed in Medline (15/17 in [2], 8/8 in [9], and 8/9 in [10]), which confirms Medline as the most efficient bibliographic database for retrieving original articles in the biomedical informatics field.

The optimized query didn't retrieve articles yet included in the reviews, 1 in [2], 3 in [9], and 5 in [10]. Taking into account that the article in [2] was also included in [9], the optimized query missed 8 articles. For all these 8 articles, the term *nurse* or its variation was absent in the title and in the abstract (e.g., more general terms were utilized

like healthcare workers). The article found in [2] and [9] should have been retrieved in the RL, and will be added to the final list of results. When refining the query (to get from 1581 returned articles to 3283), we tried to include the relevant missing articles. However, we stuck to a query we wanted to be focused on CDSSs for nursing practices taking the risk to miss some articles (mainly because of indexing errors in such articles). The remaining seven articles were out of the scope (and will not be added to RL).

Table 1. Precision and recall for the 3 selected reviews computed from papers published in [2014;2017], articles retrieved from BDBs and Medline (M), included in RL and included in RL but not retrieved by the review.

Review	#returns in BDBs/M	# articles in BDBs/M	# articles [2014-2017] in BDBs/M	# articles [2014-2017] in RL and in review	# articles [2014-2017] in RL and not in review	Precision on M	Recall on M
Akbar *et al.* [2]	1019/141	28/25	17/15(12)	11	27	91.70%	28.90%
Borum [9]	29/ NA	9/9	8/8	5	2	62.50%	71.40%
Mebrahtu *et al.* [10]	49,852/ 9,549	35/33	9/8	3	8	37.50%	27.30%

When performing a specific Medline search, we cannot foresee to what extent the search will be successful. Calculating precision and recall can help in understanding and tuning the search strategy. Therefore, when starting a systematic review, it is of major importance to have some sense of how the query is performing. Given the inverse relationship between precision and recall, unusually high precision rate may be indicative of overly restrictive searching at the expense of the complete identification of relevant studies (recall). For instance, the review of Akbar *et al.*, [2] was interested in studying the effects of CDSSs solely used by nurses on decision-making, care delivery, and patient outcomes. Despite the large scope and the number of outcomes included, it achieved a high precision and a low recall. This can be related to missing synonyms and controlled vocabulary (MeSH terms) in the query, narrowing the results to only 141 articles from Medline despite many outcomes were searched.

Having almost the same scope, Mebrahbtu *et al*. [10] studied the effects of CDSSs on nursing and *allied* health professional performance, and on patient outcomes. This review has low precision and low recall. However, the initial query for this review was made of very broad terms and the authors deliberately aimed at a high recall (49,852 returns). Reviewers must be prepared to maintain rigorous screening standards in the face of large search retrievals. It is also important to know that Mehrabtu *et al*. [10] limited the types of studies to include only methodologically sound studies (such as randomized controlled trials), which may explain why we only added 8 studies from the RL, as compared to Akbar *et al.* [2] where we added 27 articles.

Borum [9] has a very specific scope focused on the barriers to the use of CDSSs perceived by nurse practitioners in hospital settings. The high recall may explain why we didn't find a significant number of missing articles regarding this topic.

Some limitations of our work are that we only assessed Medline database, we restrained the time frame to [2014;2017], we only considered three reviews, and we restricted the search to the literature on CDSSs for nursing practice. In addition, we proposed a proxy measure for precision and recall assuming the RL was the gold

standard. Thus, true positives in our work may be different from those considered as true positives in the three reviews. Lastly, it is important to take into consideration that missing references in reviews may be due to factors beyond the control of reviews' authors. Indexing errors (coming from missing MeSH terms, or forgotten common terms in title or abstract) may lead to make relevant articles unavailable to systematic reviews.

5. Conclusion

Errors in queries are present in the majority of reviews which may limit recall and precision of search strategies. A good compromise between the two measures should be sought, in order to reach an acceptable tradeoff between result comprehensiveness and resource utilization/workload. This is why, before conducting a literature review in the domain of computerized decision support systems in nursing, we wanted to develop a comprehensive search strategy by identifying the necessary concepts, finding MeSH terms and synonyms, avoiding syntax errors, and comparing and gauging the query to the body of existing knowledge and fine tuning it in order to have a comprehensive return.

While waiting for machine learning techniques to automate the screening of relevant studies while conducting systematic reviews, human effort would be required for the foreseeable future.

References

[1] Piscotty R, Kalisch B. Nurses' use of clinical decision support: A literature review. Vol. 32, CIN - Computers Informatics Nursing. 2014. doi: 10.1097/CIN.0000000000000110.

[2] Akbar S, Lyell D, Magrabi F. Automation in nursing decision support systems: A systematic review of effects on decision making, care delivery, and patient outcomes. Vol. 28, Journal of the American Medical Informatics Association. 2021. doi: 10.1093/jamia/ocab123

[3] Salvador-Oliván JA, Marco-Cuenca G, Arquero-Avilés R. Errors in search strategies used in systematic reviews and their effects on information retrieval. Journal of the Medical Library Association. 2019;107(2). doi: 10.5195/jmla.2019.567

[4] Straube S, Heinz J, Landsvogt P, Friede T. Recall, precision, and coverage of literature searches in systematic reviews in occupational medicine: an overview of Cochrane Reviews. GMS Medizinische Informatik, Biometrie und Epidemiologie [Internet]. 2021 [cited 2023 Jan 8];17(A). Available from: https://www.egms.de/static/en/journals/mibe/2021-17/mibe000216.shtml

[5] Anderson JA, Willson P. Clinical decision support systems in nursing: Synthesis of the science for evidence-based practice. CIN - Computers Informatics Nursing.2008;26(3).doi: 10.1097/01.NCN.0000304783.72811.8e.

[6] Randell R, Mitchell N, Dowding D, Cullum N, Thompson C. Effects of computerized decision support systems on nursing performance and patient outcomes: A systematic review. Vol. 12, Journal of Health Services Research and Policy. 2007. doi: 10.1258/135581907782101543.

[7] Lee S. Features of computerized clinical decision support systems supportive of nursing practice: A literature review. Vol. 31, CIN - Computers Informatics Nursing. 2013. doi: 10.1097/01.NCN.0000432127.99644.25.

[8] Dunn Lopez K, Gephart SM, Raszewski R, Sousa V, Shehorn LE, Abraham J. Integrative review of clinical decision support for registered nurses in acute care settings. Vol. 24, Journal of the American Medical Informatics Association : JAMIA. 2017. doi: 10.1093/jamia/ocw084.

[9] Borum C. Barriers for Hospital-Based Nurse Practitioners Utilizing Clinical Decision Support Systems: A Systematic Review.Vol. 36,CIN - Computers Informatics Nursing.2018.doi: 10.1097/CIN.0000000000000413.

[10] Mebrahtu TF, Skyrme S, Randell R, Keenan AM, Bloor K, Yang H, et al. Effects of computerised clinical decision support systems (CDSS) on nursing and allied health professional performance and patient outcomes: A systematic review of experimental and observational studies. Vol. 11, BMJ Open. 2021. doi: 10.1136/bmjopen-2021-053886.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230212

Qualitative Assessment of Implementation of a Discharge Prediction Tool Using RE-AIM Framework

Joseph FINKELSTEIN[a,1], Irena PARVANOVA[a], Zhaopeng XING[a], Tuyet-Trinh TRUONG[a] and Andrew DUNN[a]
[a]*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract. The implementation process in the routine clinical care of a new predictive tool based on machine learning algorithms has been investigated using the RE-AIM framework. Semi-structured qualitative interviews have been conducted with a broad range of clinicians to elucidate potential barriers and facilitators of the implementation process across five major domains: Reach, Efficacy, Adoption, Implementation, and Maintenance. The analysis of 23 clinician interviews demonstrated a limited reach and adoption of the new tool and identified areas for improvement in implementation and maintenance. Future implementation efforts of machine learning tools should support the proactive engagement of a wide range of clinical users since the very initiation of the predictive analytics project, provide higher transparency of the underlying algorithms, employ broader onboarding of all potential users on a periodic basis, and collect feedback from clinicians on an ongoing basis.

Keywords. Clinical decision support, implementation science, machine learning

1. Introduction

Clinical decision support (CDS) tools powered by artificial intelligence and embedded into electronic health records (EHR) are considered disruptive technologies. Their integration into practice has sometimes been slow and problematic. Even considering numerous reports of the benefits of CDS, when evidence meets the realities of practice [1,2], successful deployment and adoption can be threatened. Several clinical decision support tools driven by machine learning algorithms have been recently implemented into routine clinical practice at the Mount Sinai Health System (MSHS). These algorithms were embedded into Epic EHR and made available for daily use by clinical staff caring for hospitalized patients. One of these algorithms is a 48-hr discharge prediction tool (48DPT) which allows forecasting discharge dates for hospitalized patients based on discrete clinical data available in the EHR. The 48DPT is currently being used by hospitalist service during daily interdisciplinary rounds. However, the actual uptake of the 48DPT, its acceptance by clinical staff for routine clinical care, and its perceived and documented impact on care quality are unknown. Implementation science posits RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework as a validated means to assess the implementation of complex

[1]Corresponding Author: Joseph Finkelstein, E-mail: Joseph.Finkelstein@mssm.edu

interventions in health care [3]. The goal of this project was to conduct a systematic evidence-based assessment of the implementation of the 48-hr Discharge Prediction Tool using the RE-AIM Framework.

2. Methods

A purposive sample comprised healthcare providers at MSHS (e.g., attending faculty physicians, residents, advanced practice nurses, staff nurses, social workers, and hospitalists). We attempted to include a sample with a wide range of expertise and years of experience in hospital care. At the beginning of each interview, verbal consent was collected. Each interview took approximately 30 minutes and was professionally transcribed verbatim. To instruct the interview process, a moderator guide was developed based on five dimensions of the RE-AIM framework [3]. A thematic analysis was conducted to identify, analyze, and report the themes that emerged from the participant's responses. The analysis was guided by the RE-AIM framework and followed the six-step process proposed by Braun and Clarke [4]. The study was approved by the MSHS IRB.

3. Results

The semi-structured interviews were conducted on three medical units with 23 participants, including sixteen hospitalists, four of whom are unit medical directors (UMD), four social workers (SW), two case managers (CM), and one nurse manager (NM). The analysis ended up with 120 codes and 25 themes. The themes describe the current situation and rising challenges of the 48DPT implementation. Four of the themes (providing training, increasing awareness, improving integration, and improving accuracy) demonstrate the participants' most frequently mentioned resolutions and suggestions to improve the long-term usage of 48DPT. In the following framework, each theme was presented with the number of unique participants who mentioned it.

The **Reach dimension** of this analysis focused on the targeted users who were aware of and used the 48DPT. Four major themes were identified, i.e., awareness, use frequency, user cohorts, and patient representativeness. *Awareness:* This theme indicates if a participant had heard of 48DPT. 14 participants reported that they were aware of this tool, including two UMDs, two CMs, two SWs, and 8 attendings, mostly affiliated with a specific medical unit at Mount Sinai Hospital (KCC unit); whereas 5 reported that they had never heard of this tool. *Use frequency:* This theme indicates if a participant was a frequent user. Out of the 14 participants who were aware of 48DPT, 7 reported frequent use at a team level, including 2 CM, 2 SW, 1 UMD, and 2 hospitalists (i.e. frequent users). Among these frequent users, 6 reported a daily frequency. In comparison, the other 16 participants, 9 of whom were hospitalists, reported a rare or no use of 48DPT at either team or individual level (i.e., no/rare user). *User cohorts:* This theme reflects participants' perception of frequent and non-frequent users of 48DPT. 12 participants perceived that 48DPT had potential benefits and was specifically used by a few types of roles, including CMs, SWs, and unit medical directors. In comparison, 3 participants stressed that 48DPT was not their focus on everyday routine (i.e., an NM, a SW, and a hospitalist). *Patient's representativeness:* This theme corresponds to questions regarding the use of 48DPT varied by patient-related factors. 14 participants observed no differences in using 48DPT

in terms of patients' related factors, whereas 5 reported differences in terms of patients' age, and disease severity.

The **Efficacy dimension** measures the impacts of 48DPT on the predefined outcomes, i.e., prediction accuracy, length of stay (LOS), initiating discharge process, and clinical patient outcomes. In addition, it also covers the participants' perceptions of such impacts in comparison with humans and Discharge Today (DT). *Accuracy (Medical readiness):* Because many participants used "accuracy" to explain the predefined outcome "effectiveness in identifying medical readiness", we thus used the accuracy to describe the codes related to such outcome. A participant reported that the prediction of 48DPT was accurate, and another participant showed trust in the 48DPT result. There were 5 participants who gave a range of estimated accuracy of 60-80%. In comparison, 6 participants reported that 48DPT was not consistently accurate; 3 participants reported that follow-up human validations were necessary. *LOS:* The theme investigates the impacts of 48DPT on the shortening of the patient's length of stay. 9 participants reported that the use of 48DPT didn't decrease LOS, and none of the participants stated that 48DPT did decrease LOS. *Initiating the discharge process:* This theme represents participants' perceptions of the efficacy of 48DPT in initiating a discharge process. 10 participants reported that 48DPT was helpful in initiating a discharge process. *Clinical care/patient outcomes:* This theme represents participants' perceptions of the efficacy of 48DPT in improving clinical care or patient outcomes. 6 participants considered that 48DPT had improved the outcomes. *Comparison with Humans:* This theme demonstrated the comparison made between 48DPT with the human judgment of discharge readiness. 3 participants reported that 48DPT was more or equally accurate than human judgment, and another two observed that the results were consistent with human judgment. In comparison, 10 participants reported that 48DPT was less accurate than humans or that the prediction was inconsistent with human judgment. 5 participants reported a tendency to trust humans instead of 48DPT, and only 1 participant showed trust in the prediction results of 48DPT. *Comparison with DT:* This theme indicates the comparison made between 48DPT with DT. There were 21 participants who were familiar with and used DT, and four of them showed a strong preference for using DT to 48DPT. 3 participants stressed that they used DT because their teams used it; it was multidisciplinary and team-coordinated. When it came to functionality, 5 participants recognized the differences between the 48DPT and DT. 3 participants appreciated its automated prediction function, which released them from manual data input. 8 considered the two to be functionally redundant. In addition, 5 participants considered DT to be more useful and accurate. 9 participants pointed out that DT considered patients' social factors in a discharge process, and 4 added that DT provided context information about a discharge of a patient, lacking in 48DPT.

The **Adoption dimension** addressed questions regarding the practices of using 48DPT at individual and unit levels and user acceptability in terms of different specialties and levels of experience in clinics. *Intended usage:* This theme reflects the participant's thoughts if 48DPT was used as intended. 7 confirmed that the 48DPT was used as intended, but 3 disagreed. *Usage at the unit level:* This theme captures participants' thoughts regarding the variety of 48DPT usage across different units. 6 participants reported that the use of 48DPT varied across the unit. The difference primarily lies in the use frequency. *Acceptability at an individual level:* This theme shows participants' reactions and responses to the use of 48DPT in the patient discharge workflow. 4 participants thought that the attitudes were varied in terms of physician's roles, whereas 3 didn't think there were differences. Also, 2 observed that the attitudes were varied in

terms of time in clinics where the younger practitioners would be more open to adopting the use of 48DPT. 4 didn't observe any variations. In addition, 2 clinicians showed a willingness to use 48DPT and perceived it to be useful in facilitating the patient discharge process.

The **Implementation dimension** focuses on the predisposing and enabling factors for a successful implementation. In this analysis, 2 themes were associated with the predisposing factors, i.e., perceived usefulness and easiness to use, computer skills, while 4 were associated with the enabling factors, i.e., involvement in the development process, integration into existing workflow, system transparency, and user training. *Perceived usefulness and easiness-to-use (Predisposing):* This theme represents participants' perceptions of the utility and easiness-to-use of 48DPT. 7 thought 48DPT to be useful. Some believed it would be especially useful for practitioners, SW, and sicker patients. 3 thought that 48DPT was accessible and intuitive to use. In contrast, 2 participants thought that 48DPT was not useful, and 5 were unsure. *Computer skills (Predisposing):* This theme reflects participants' thoughts about technical proficiency for 48DPT. 12 participants reported that 48DPT required basic computer skills. 2 reported that it required intermediate skills. *Involvement in the development process (Enabling):* This theme addresses participants' observations of involvement in the development process at both individual and team levels. Only 1 UMD reported that he/she was involved in the development process, whereas 18 participants were not involved. 6 participants reported the interdisciplinary team was involved in the development process, and 3 reported that the team wasn't involved. *Integration into existing workflow (Enabling):* This theme addressed the integration of 48DPT into the existing patient discharge workflow. 4 participants reported that the tool was well-integrated, and 1 reported it was somewhat integrated, whereas 11 participants disagreed that the tool was integrated. *Transparency (Enabling):* This theme describes the existing challenges caused by the lack of system transparency to the participants. 10 participants reported a lack of knowledge about how 48DPT works and its accuracy in predicting medical readiness. *Training (Enabling):* This theme covers the codes related to the shortage of training provided for users on using 48DPT. 12 participants reported a lack of training on how to access, use, and integrate 48DPT. The major concern reported by 9 participants was not being able to interpret the result. Another 5 reported uncertainty about which training to take and how it should be provided. 1 SW reported no training needed.

The **Maintenance dimension** addresses the reinforcement factors for 48DPT to be part of IDR and discharge procedure at Mount Sinai. There were six reinforcement factors derived from codes, burden vs. benefits, continuous support, increasing awareness, improving accuracy, improving integration, and providing training. *Burden vs. Benefits:* This theme shows patients' perceptions of the burdens versus benefits of using 48DPT. 8 participants reported no burden when using the 48DPT. 6 participants considered the use of 48DPT as a type of additional burden which was primarily due to the complexity of the existing patient discharge workflow. 9 participants recognized the benefits of using 48DPT while another 4 did not. *Continuous support:* This theme summarizes the codes regarding the long-term efforts made by the leadership and development team in continuously supporting the use of 48DPT. 11 participants reported no efforts observed from the leadership to receive feedback on the use of 48DPT from users, while 4 reported that the feedback requests were collected in different ways. 9 participants reported that no changes had been made to 48DPT over time; 12 participants reported no effort to improve the use of the 48DPT. 1 UMD reported a lack of follow-up analysis to understand the use of the tool in clinics. *Increasing awareness:* This theme

reflects the participants' suggestions for improving the use of 48DPT by increasing awareness. 11 participants proposed that the team should increase the awareness of the rollout of the 48DPT, especially among the hospitalists, NMS, ID team, and SW. 3 suggested the importance of leadership in promoting the utility of the tool. 2 highlighted the importance of making the users aware of the purpose and utility of the tool. A UMD suggests the consideration of the impacts of rapid employment turnover rate on the dissemination and utility of 48DPT. *Improving accuracy:* This theme reflects the participants' suggestions for improving the use of 48DPT by improving the accuracy of prediction. 6 participants suggested the importance of increasing accuracy for improving the maintenance, and 2 suggested resolving the inconsistency between 48DPT with humans and DT. *Improving integration:* This theme reflects the participants' suggestion for improving the use of 48DPT by facilitating system integration. 3 participants suggested the integration of 48DPT into existing IDR, and another 3 suggested the integration with DT. To optimize the integration, 3 participants suggested considering patients' social factors in the design of 48DPT. 1 participant stressed that 48DPT should simplify and streamline the existing complex workflow instead of imposing additional burdens. *Providing training:* This theme reflects the participants' suggestions for improving the use of 48DPT by providing training sources. 15 participants proposed to receive training regarding the use, access, integration, and interpretation of the 48DPT, which was especially important for frequent users, such as CM, SW, UMD, residents, and hospitalists.

4. Discussion and Conclusion

The analysis of 23 clinician interviews demonstrated a limited reach and adoption of the new tool and identified areas for improvement in implementation and maintenance. Our results concur with previous studies that identified benefits to patients, low-level technical skill requirements, and intended usage as some of the main factors leading to the high acceptance of a clinical decision support tool [5,6]. Lack of accuracy and awareness, as well as lack of accessibility and level of transparency, led to the low acceptance of 48DPT. Future implementation efforts of machine learning tools should support the proactive engagement of a wide range of clinical users since the very initiation of the predictive analytics project, provide higher transparency of the underlying algorithms, employ broader onboarding of all potential users on a periodic basis, and collect feedback from clinicians on an ongoing basis.

References

[1] Han Y, Carcillo J, Venkataraman S, et al. Unexpected mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics. 2005;116(6):1506-12.
[2] Cha E, Golden SH, Finkelstein J. Unanticipated consequences of hospital-based insulin management order sets. Stud Health Technol Inform. 2015;216:939.
[3] Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM Planning and Evaluation Framework: Adapting to New Science and Practice with a 20-Year Review. Front Public Health. 2019;7:64.
[4] Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.
[5] Bedra M, Hill Golder S, Cha E, et al. Computerized Insulin Order Sets Can Lead to Unanticipated Consequences. Stud Health Technol Inform. 2015;213:53-6.
[6] Kawamoto K, Finkelstein J, Del Fiol G. Implementing Machine Learning in the Electronic Health Record: Checklist of Essential Considerations. Mayo Clin Proc. 2023;98(3):366-369.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230213

Decision Support for Signal Assessment of Large Case Series in Pharmacovigilance

Lucy QUIRANT[a], Lovisa SANDBERG[a], Jim W BARRETT[a,1] and Johan ELLENIUS[a]
[a] *Uppsala Monitoring Centre, Uppsala, Sweden*

Abstract. In pharmacovigilance, signal assessment of a medicinal product and adverse event can involve reviewing prohibitively large numbers of case reports. A prototype of a decision support tool guided by a needs assessment was developed to help manual review of many reports. In a preliminary qualitative evaluation, users said the tool was easy to use, improved efficiency and provided new insights.

Keywords. Decision support, pharmacovigilance, signal assessment

1. Introduction

Post marketing safety surveillance involves monitoring medicinal products already in routine clinical use for safety concerns not detected during the pre-marketing clinical trials. At Uppsala Monitoring Centre (UMC), this involves screening VigiBase, the WHO global database of individual case safety reports, which contains over 33 million reports. A signal is a hypothesis of "a new potentially causal association, or a new aspect of a known association, between an intervention and an event" [1].

Signal assessors (SA) investigate case series which can contain hundreds or even thousands of reports about a medicinal product and adverse event. As the size of case series increases, so does the complexity of signal assessment. Currently, all reports are manually reviewed, a task that can be prohibitively time consuming. We hypothesized that one way of addressing this problem is by providing a set of decision support functionalities that allow SAs to identify subsets of reports for analysis from specific points of view. Other groups have proposed similar hypotheses [2]. To this end, we carried out a needs assessment, built a tool prototype, and evaluated its usefulness and usability.

2. Methods

Six SAs at UMC were interviewed to capture their needs regarding assessment of large case series. From the interviews we derived a common workflow for the signal assessment process, and its associated challenges. One of the needs identified was in identifying reports relating to specific risk factors or confounders, as finding such patterns is highly time consuming.

[1] Corresponding Author: Jim W Barrett, E-mail: jim.barrett@who-umc.org

A decision support tool was built using Python and the Streamlit framework to facilitate the users' needs, including the definition of risk factors through coded (MedDRA) medical and free text search terms within the relevant fields in reports. Predefined risk factor definitions were provided to identify cases involving patients with COVID-19 disease, and tools were implemented allowing users to define their own risk factors. Defined risk factors can be stored in a knowledge repository, for later reuse in the analysis of other case series. This allows users to explore salient subsets of reports.

A preliminary assessment was conducted on the effectiveness of the predefined risk factors. A small-scale qualitative evaluation was conducted to assess usability and usefulness. Four SAs used the tool for a week and received a questionnaire which was followed by a focus group discussion to capture additional feedback.

3. Results

In a preliminary experiment on a single vaccine adverse event combination with 123 reports, the predefined risk factor had a recall of 0.81 and a precision of 0.59 in finding the 52 true relevant reports. The outcome of the questionnaire and focus group discussion suggest that the tool is easy to use and effective in reducing the time required to perform an analysis. SAs said they would use the current version as part of their signal assessment process. The tool allowed users to find information in reports faster, confirm previous manual findings and detect new relevant reports. Possibilities to explore ideas more easily gave new insights into the case series. Features for collaboration between assessors and further editability were wished for.

4. Discussion

A caveat in this qualitative evaluation was that users only used the tool in the late stages of signal assessment. In the future, a more in-depth evaluation is planned when users have used the tool for longer, including in the early phases of signal assessment.

5. Conclusions

This preliminary study indicates that a decision support tool that enables the identification of subsets of case reports to reduce the complexity of pharmacovigilance signal assessment is considered useful by SAs and has the potential to improve their time efficiency.

References

[1] Council for International Organisations of Medical Sciences (CIOMS) Working Group VIII. Practical aspects of signal detection in pharmacovigilance CIOMS; 2010.

[2] Information Visualization Platform for Postmarket Surveillance Decision Support. Drug Saf. 2020 Sep;43(9):905-915. doi: 10.1007/s40264-020-00945-0. PMID: 32445187.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230214

Explainable Graph Neural Networks for Atherosclerotic Cardiovascular Disease

Jens LUNDSTRÖM[a,1], Atiye Sadat HASHEMI [a] and Prayag TIWARI [a]
[a] *Center for Applied Intelligent Systems Research in Health, Halmstad University, Sweden*

Abstract. Understanding the aspects of progression for atherosclerotic cardiovascular disease and treatment is key to building reliable clinical decision-support systems. To promote system trust, one step is to make the machine learning models (used by the decision support systems) explainable for clinicians, developers, and researchers. Recently, working with longitudinal clinical trajectories using Graph Neural Networks (GNNs) has attracted attention among machine learning researchers. Although GNNs are seen as black-box methods, promising explainable AI (XAI) methods for GNNs have lately been proposed. In this paper, which describes initial project stages, we aim at utilizing GNNs for modeling, predicting, and exploring the model explainability of the low-density lipoprotein cholesterol level in long-term atherosclerotic cardiovascular disease progression and treatment.

Keywords. Graph Neural Networks, Cardiovascular Diseases, EHR

1. Introduction and Methods

Atherosclerotic cardiovascular disease (ASCVD) [2] remains one of the leading causes of mortality in the world. Despite significant improvement in ASCVD treatments, the burden of ASCVD risk factors remains high and poses a significant economic burden. High levels of low-density lipoprotein cholesterol (LDL-C) have been proven to be a key risk factor for ASCVD and the treatment of dyslipidemia represents a vital strategy to reduce new cardiac events as well as mortality [1]. Secondary prevention strategies are crucial in order to reduce recurrence and decrease morbidity and mortality. In this paper, which describes the initial stages of our newly started project, we aim to use real Electronic Health Records (EHR) to model and interpret the disease progression of ASCVD after the index event (e.g., stroke or myocardial infarction) where the main target signal is the reduction of LDL-C. To interpret key factors of LDL-C reduction we also aim to leverage post-hoc XAI [2] techniques for Graph Neural Networks (GNN) to predict and monitor the long-term treatments (0-5 years after the index event) in ASCVD. The task of modeling disease progression is challenging due to several reasons, among the challenges: different modalities of EHR, temporal and high dimensional patient trajectory data, episodic and irregular frequency of hospital visits [3]. The data to be processed contains a graph-like temporal structure of hospital visits linking to patient demographic, diagnosis codes, and lab test results.

[1] Corresponding Author: Jens Lundström, E-mail: jens.r.lundstrom@hh.se.

[2] Build-up of fatty deposits in the inner walls of blood vessels (arteries) causing restricted or blocked blood flow.

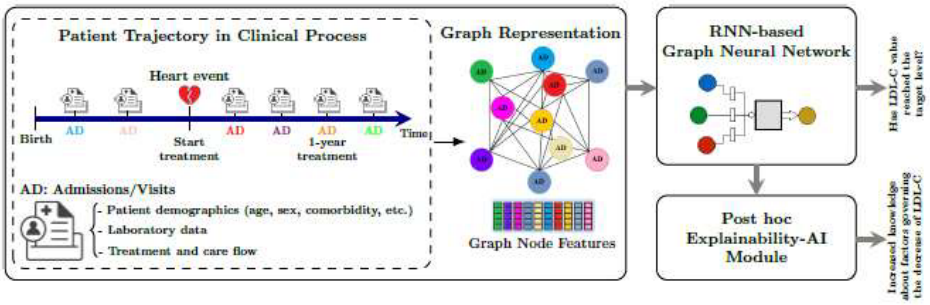


Figure 1. Our proposed framework.

Figure 1 shows a recurrent neural network (RNN)- based GNN for predicting and monitoring the ASCVD progression and treatment. Each node is an admission and the graph could be considered as evolving over time which makes methods such as Temporal Graph Networks suitable. Our network could potentially be the engine of a decision support system guiding the clinician to a patient-specific treatment to maximize the likelihood to reach the LDL-C target. Moreover, the XAI module allows the framework to obtain knowledge about factors governing the LDL-C level changes.

2. Results, Discussion and Conclusion

One of the most interesting aspects of working with EHRs in this project is how high-risk ASCVD patients' trajectories could be described from a demographic, clinical, and resource perspective. GNNs show promising abilities for modeling the complex relations of these graph-structured EHRs. Recent works focus on GNN explainability but it remains an important topic to explore further, especially in the healthcare domain. Existing approaches for GNN XAI leverage surrogate models, perturbation-based explanation, and gradient-based methods [2]. However, it is our belief that the development of explainable decisions from a decision support system of ASCVD requires not only technical methods transforming embeddings of EHRs to interpretable information but that such methods need to be well-aligned with clinical practice and knowledge. These multidisciplinary aspects are the core of this study, which is continued in the future work along with approved ethical application and data access. Making use of XAI for cardiovascular diseases is nothing new, however to the best of our knowledge the project focus on explainable GNN-based prediction of ASCVD progression is novel.

References

[1] Sundbøll J, Larsen AP, Veres K, Adelborg K, Sørensen HT. Cardiovascular event rates and trajectories of LDL-cholesterol levels and lipid-lowering therapy in patients with atherosclerotic cardiovascular disease: a population-based cohort study. Thrombosis Research. 2019 Nov 1;183:124-30.

[2] Feng Q, Liu N, Yang F, Tang R, Du M, Hu X. DEGREE: Decomposition Based Explanation for Graph Neural Networks. InInternational Conference on Learning Representations 2022.

[3] Tang S, Tariq A, Dunnmon JA, Sharma U, Elugunti P, Rubin DL, Patel BN, Banerjee I. Predicting 30-day all-cause hospital readmission using multimodal spatiotemporal graph neural networks. IEEE Journal of Biomedical and Health Informatics. 2023 Jan 13.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230215

Guideline-Based Algorithmic Recommendations Versus Multidisciplinary Team Advice for Gynecologic Oncology

Kees EBBEN[a,1], Marleen van HOUDT[a], Cor de KROON[b] and Jurrian van der WERF[a]
[a] *Department of Research and Development, Netherlands Comprehensive Cancer Organization, Utrecht, The Netherlands*
[b] *Department of Gynecology and Obstetrics, Leiden University Medical Center, Leiden, The Netherlands*
ORCiD ID: Kees C.W.J. Ebben https://orcid.org/0000-0003-1069-2357

Evidence-based clinical decision making in oncology is challenging. Multidisciplinary team (MDTs) meetings are organized to consider different diagnostic and treatment options. MDT advice are often based on clinical practice guideline recommendations which can be extensive and ambiguous, making it difficult to implement in clinical practice. To address this issue, guideline-based algorithms have been developed. These are applicable in clinical practice and enable accurate guideline adherence evaluation. This ongoing study aims to determine the optimal decision-making approach for different subpopulations of patients with high-incidence gynecological cancers.

Keywords. Oncology, Multidisciplinary team, Decision Making, Clinical guideline

1. Introduction

Evidence-based clinical decision-making in oncology is increasingly challenging due to numerous subpopulations and the large variety of available treatment options [1]. Multidisciplinary team (MDTs) discussions serve to obtain insight regarding patient and disease characteristics on an aggregated level, to consider diagnostic and treatment options and to reach a multidisciplinary advice based on clinical practice guidelines recommendations [2]. Recommendations in textual guidelines are often extensive, ambiguous and inconsistent and not systematically aligned with the clinical decision process in the care path. A method that remodels guideline recommendations into unambiguous, data-driven decision algorithms is clinical decision trees (CDTs) [3]. To manage relevant characteristics for making guideline-based recommendations MDTs could potentially benefit from a computerized clinical decision support system (CDSS). Little attention has been paid to the role of CDSS preceding a MDT, for selection of cases that benefit from a multidisciplinary discussion. It is expected that subpopulations can be identified who do not benefit from a multidisciplinary discussion [1]. The aim of this paper is to present a currently ongoing study exploring the research question: For

[1] Corresponding author: Kees C.W.J. Ebben, Hatertseweg 1, 6533 AA, Nijmegen, the Netherlands, E-mail: k.ebben@iknl.nl.

every subpopulation of the total population under study, which decision-making echelon ('guideline-based algorithmic' versus 'multidisciplinary team') achieves the optimum balance between quality of care for the patient and physicians' efforts?

2. Methods and Outcomes

This prospective, multicenter, observational, concordance study takes place at 21 hospitals, 5 gynecologic oncologic centers and their network clinics. We aim to include a total of 300 cases, distributed over 15 CDTs, associated with 3 gynecological oncological diseases (cervical cancer, endometrial cancer and epithelial ovarian cancer). A senior medical doctor observes MDT discussions and manually collects all available data during MDTs. After each MDT data from all cases are plotted onto the corresponding CDT in order to generate a guideline-based recommendation. The proposed interventions from MDT and CDT are assigned to one of the four following groups, depending on the level of concordance: I) concordant, II) conditionally concordant, III) non-concordant (motivated) and IV) non-concordant (non-motivated).

Based on concordance levels, performed treatment and achieved health outcomes a proposal for echelons in gynecologic oncological decision-making is made for each included subpopulation. The following echelons could be considered: 1. Monodisciplinary with guidelines (algorithmic), 2. Multidisciplinary (local hospital), 3. Multidisciplinary with experts (consultation of gynecologic oncological center).

3. Discussion and Conclusion

Expected results of this study will provide guidance for defining appropriate decision-making levels for gynecological oncology. In the Netherlands, healthcare professionals have set standards for discussing at least 90% of cancer patients This research may help to nuance current standards [8]. Results are determined at population level, meaning that for an individual patient the optimal level of decision-making may differ from the rest of her subpopulation. For this reason, we recommend that the treating physician always remains in control to determine the optimal decision-making echelon for each individual patient. Conclusion: With the help of the proposed applied health informatics innovation, the pressure on care professionals in oncology may be reduced, while quality of care is maintained. Preliminary results of the study are expected in the second half of 2023.

4. References

[1] Glatzer M, Panje CM, Sirén C, Cihoric N, Putora PM, Decision Making Criteria in Oncology, 98 (2020), 370-378, doi: 10.1159/000492272.

[2] Pillay B, Wootten AC, Crowe H, et. al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: A systematic review of the literature, Cancer Treat Rev 42 (2016), 56-72, doi: 10.1016/j.ctrv.2015.11.007.

[3] Hendriks MP, Verbeek XA, van Vegchel T, et. al. Transformation of the National Breast Cancer Guideline Into Data-Driven Clinical Decision Trees, JCO Clin Cancer Inform 3 (2019), 1-14, doi: 10.1200/CCI.18.00150.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230216

Diagnosis Support for Rare Diseases Using Phenotypic Profiles

Richard NOLL[a,1], Jannik SCHAAF[a], Michael VON WAGNER[b] and Holger STORF[a]
[a] *Goethe University Frankfurt, University Hospital Frankfurt, Institute of Medical Informatics, Frankfurt, Germany*
[b] *Department of Medical Information Systems and Digitalization, University Hospital Frankfurt, Frankfurt, Germany*

Abstract. The common occurrence of characteristic symptoms can be used to infer diagnoses. The aim of this study is to show how syndrome similarity analysis using given phenotypic profiles can help in the diagnosis of rare diseases. HPO was used to map syndromes and phenotypic profiles. The system architecture described is planned to be implemented in a clinical decision support system for unclear diseases.

Keywords. Clinical Decision Support, Rare Disease, Syndrome

1. Introduction

The common occurrence of certain characteristic symptoms, also known as a syndrome, can be used to infer diagnoses. The diagnoses give affected individuals and their families access to resources, prognosis, and available treatments. There is an increasing use of expert systems to support syndrome diagnosis, including computer databases and analytic software, but also human expert and online services [1].

The aim of this study is to show how syndrome similarity analysis using given phenotypic profiles containing observable characteristics can help diagnose rare diseases (RD). This study describes a feasible system architecture to be used as an interoperable module of a case-based clinical decision support system (CDSS) for unclear diseases.

2. Methods

First, such a system must capture symptoms in a standardized form ensuring accurate similarity analysis. One ontology that covers a wide range of bioinformatics resources for human disease analysis, especially RDs, is the Human Phenotype Ontology (HPO) [2]. Complete phenotypic profiles on RDs can be accessed via Monarch [3] or HPO [2].

The selection of the similarity algorithm is crucial for the quality of the analysis. For HPO, there is the Python library *pyhpo* [4], which is open source and offers the possibility to define and compare different symptom sets. This functionality enables the module to compare syndromes of a new patient with given phenotypic profiles of RDs.

The usage of the HL7 data exchange standard FHIR ensures interoperability.

[1] Corresponding Author: Richard Noll, Institute of Medical Informatics, Goethe University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; E-mail: richard.noll@kgu.de.

3. Results

Figure 1 shows a system architecture that determines the similarity of the syndrome of a new patient case with the phenotypic profiles of given RDs. In addition to the calculated percentage similarity, further information such as the indication of identical, similar, or missing symptoms to the different phenotypic profiles is conceivable as an output.

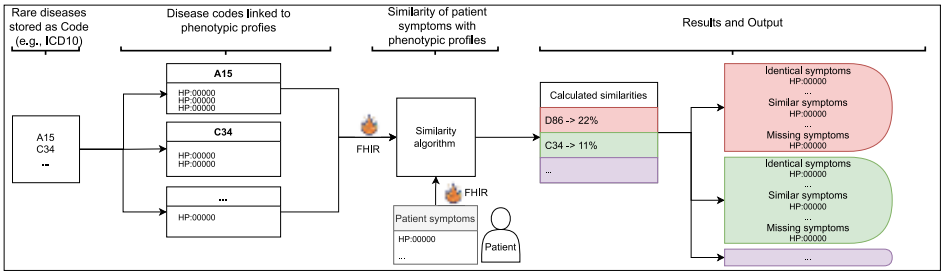


Figure 1. System architecture of an interoperable module providing disease prognosis based on a syndrome.

4. Discussion

The applied similarity algorithm works similarly to the *Phenomizer* [5], a clinical diagnostic web application that uses semantic similarity searches in ontologies, which however has no API. This study shows such an algorithm in conjunction with FHIR.

The advantage of an interoperable module over an isolated web-based platform is the cross-system operability and the possibility of a combined analysis of symptoms along with additional characteristics such as demographic data or laboratory tests to accumulate distinctive information and make an accurate differential diagnosis.

5. Conclusion

To validate the feasibility, the next step is to implement the described system architecture as an interoperable module into SATURN [6], a project funded by the German Federal Ministry of Health which aims to develop a CDSS for patients with unclear diseases.

References

[1] Hallgrímsson B, Aponte JD, […], Klein OD. Automated syndrome diagnosis by three-dimensional facial imaging. Genet Med. 2020 Oct;22(10):1682-1693. doi: 10.1038/s41436-020-0845-y

[2] Köhler S, Carmody L, […], Robinson PN. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019 Jan 8;47(D1):D1018-D1027. doi: 10.1093/nar/gky1105

[3] Mungall CJ, McMurry JA, […], Haendel MA. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2017 Jan 4;45(D1):D712-D722. doi: 10.1093/nar/gkw1128

[4] Marcello J, Pyhpo, https://github.com/Centogene/pyhpo, last access: 30.12.2022.

[5] Köhler S. Vom Symptom zum Syndrom mit moderner Softwareunterstützung [From symptom to syndrome using modern software support]. Internist (Berl). 2018 Aug;59(8):766-775. German. doi: 10.1007/s00108-018-0456-8

[6] SATURN, https://www.saturn-projekt.de/, last access: 30.12.2022.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230217

A Masked Language Model for Multi-Source EHR Trajectories Contextual Representation Learning

Ali AMIRAHMADI[a,1], Mattias OHLSSON[a,b], Kobra ETMINANI[a], Olle MELANDER[c] and Jonas BJÖRK[d]
[a] *Center for Applied Intelligent Systems Research, Halmstad University, Sweden*
[b] *Centre for Environmental and Climate Science, Lund University, Sweden*
[c] *Division of Occupational and Environmental Medicine, Lund University, Sweden*
[d] *Department of Clinical Sciences, Lund University, Sweden*
ORCiD ID: Ali Amirahmadi https://orcid.org/0000-0002-1999-8435

Abstract. Using electronic health records data and machine learning to guide future decisions needs to address challenges, including 1) long/short-term dependencies and 2) interactions between diseases and interventions. Bidirectional transformers have effectively addressed the first challenge. Here we tackled the latter challenge by masking one source (e.g., ICD10 codes) and training the transformer to predict it using other sources (e.g., ATC codes).

Keywords. representation learning, patient trajectories, Masked language model, electronic health records, deep learning, disease prediction

1. Introduction

Electronic health records (EHRs) together with advanced machine learning, have provided opportunities for the next generation of medical decision support systems. The longitudinal collection of EHR data (aka. patient trajectories) has successfully been analyzed using recent developments in natural language processing methods. Pre-trained language representations with self-supervised methods have found their counterpart for EHR data, such as EHR2Vec and EHR contextual learning [1]. Language models such as Bert and their EHR peers like Behrt [2] & MedBert [3], because of their capability to handle time dependencies have outperformed the state-of-the-art methods for predicting medical outcomes. In this work we addressed the challenge of interactions between different data sources using new masking methods for Masked language models learning.

2. Material and methods

This paper used the history of diagnoses (1.5M ICD10 codes), medications (6M ATC codes), and two questionnaires (baseline and five year follow-up with 1895 questions in total) of approximately 30,000 persons from the Malmö Diet and Cancer Cohort [4].

[1] Corresponding Author: Ali Amirahmadi, E-mail: ali.amirahmadi@hh.se

We devised a two-step masking process to learn the effective representation of the multi-source EHR data and address mentioned challenges. Firstly, to handle the long/short-term dependencies, we randomly mask some proportion of the patient trajectories and train our network to predict the masked part via the unmasked ones. Then, to model the interactions between life habits, interventions, and diseases, we mask one of the sources for a specific period and train the network to predict it using other sources during that period.

Then to learn the representation, we train a transformer encoder to predict the masked part of the input sequence data **(Figure 1)**. In the last step, we add a classifier layer on top of the trained network and fine-tune it for downstream tasks.

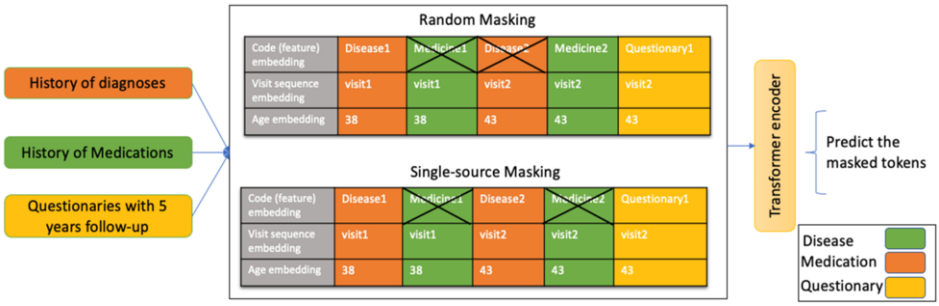


Figure 1. Learning the effective representation of multi-source patient trajectories by a two-step Masking MLM and a multi-head transformer encoder

3. Results, Discussion and Conclusion

To evaluate the built representation learning model, we fine-tuned it to predict the Heart-failure ICD10 codes in the next visit and classify the early (premature) death. Preliminary results show the promising ability of two-step MLM, especially in Few-Shot learning. Here we introduced a two-step masking process for patient trajectories representation learning. Developing an effective ML model strongly depends on efficient feature extraction from all available sources. We expect the developed model to effectively address the time dependencies and the interactions between disease, interventions, and other sources simultaneously. Incorporating data sources with other modalities like medical images still needs more investigation.

References

[1] Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. In: Proc. Machine Learning for Healthcare Conference. PMLR; 2016. p. 301–18.
[2] Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: Transformer for Electronic Health Records. Sci Rep. 2020 Apr 28;10(1):7155.
[3] Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. Npj Digit Med. 2021 May 20;4(1):1–13.
[4] Berglund G, Elmståhl S, Janzon L, Larsson SA. Design and feasibility. J Intern Med. 1993;233(1):45–51.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230218

Modeling Clinical Guidelines for an Epilepsy-CDSS: The EDiTh Project

Ariadna PÉREZ GARRIGA[a,1] Stefan WOLKING[b], Jonas FORTMANN[a], Raphael W MAJEED[a], Christian STOCKEM[a], Lukas NIEKRENZ[a], Christian BOSSELMANN[c], Yvonne WEBER[b,c], Rainer RÖHRIG[a] and Myriam LIPPRANDT[a]

[a] *Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*

[b] *Department of Epileptology and Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany*

[c] *Department of Neurology and Epileptology, Hertie-Institute for Clinical Brain Research, Eberhard-Karls University of Tübingen, Tübingen, Germany*

ORCiD ID: Ariadna Pérez Garriga https://orcid.org/0000-0001-8840-1770

Abstract. The knowledge transformation process involves the guideline for the diagnosis and therapy of epilepsy to an executable and computable knowledge base that serves as the basis for a decision-support system. We present a transparent knowledge representation model which facilitates technical implementation and verification. Knowledge is represented in a plain table, used in the frontend code of the software where simple reasoning is performed. The simple structure is sufficient and comprehensible also for non-technical persons (i.e., clinicians).

Keywords. Clinical Decision Support Systems, Computer-Assisted Decision Making, Knowledge Representation (Computer), Practice Guidelines as Topic

1. Introduction

Guideline-based Clinical Decision Support Systems (CDSS) use computer-interpretable guidelines (CIGs) to improve guideline-adherence and thus also quality of care [1]. The challenge is to transform medical knowledge into a comprehensible, modifiable CIG, testable by clinicians. This work aims to report a modular rule-based approach and the lessons learned from the EDiTh-Project. The purpose of the EDiTh-App is to support clinicians in seizure anamnesis, diagnostics, diagnosis, and therapy of epilepsy to improve guideline adherence.

2. Methods

With a group of international experts in Epilepsy, we identified two relevant guidelines: One categorizes the seizure type (ILAE) [2], and the other recommends the right therapy based on a classification of Epilepsy [3], updated by members of the guideline group.

[1] Corresponding Author: Ariadna Pérez Garriga, E-mail: aperezgarrig@ukaachen.de

The gap between seizure classification (ILEA) and the classification of various kinds of epilepsy was closed by interviews with experts and literature. The knowledge was separated into 3 models (see Figure 1): To find the right seizures type, a decision tree was chosen; to find the right diagnoses based on the seizure anamnesis and the diagnostic findings, we use a decision table; we also use a decision table for the representation of the therapy, based on the diagnosis and other individual factors (age, sex).

Figure 1. Stages of the decision process of the EDiTh-App

3. Results

The decision tree and the decision tables are implemented as configuration CSV-files for the CDSS. The user-interface (a VUE.js web-application) and the inference engine (JavaScript CSV-parser) use both these configuration tables. The user is guided through the seizure anamnesis, document the diagnostic findings, confirm the diagnosis, and get the recommended therapy. All the user-entered data and the output of each step of the CDSS is then saved in a database (PostgreSQL, JAVA backend) to reuse and evaluate it.

4. Discussion and Conclusion

The proposed modeling strategy facilitates the creation of CIG by using a simple structure for the operationalization of clinical knowledge. Clinicians can directly participate in the knowledge engineering process, and the resulting operationalization is simple to validate. This eases the technical implementation, as proved by the EDiTh-Project. The CDSS will be evaluated in a clinical study under art. 82 MDR and for the evaluation of the process of knowledge modelling, further research is needed.

References

[1] Murphy EV. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. Yale J Biol Med. 2014 Jun 1;87(2):187–97.

[2] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017 Apr;58(4):512–21.

[3] Elger C. E., Berkenfeld R. et al. S1-Leitlinie Erster epileptischer Anfall und Epilepsien im Erwachsenenalter. 2017. In: Deutsche Gesellschaft für Neurologie, Hrsg. Leitlinien für Diagnostik und Therapie in der Neurologie. [Internet] [cited 2022 Dec 16] Available from: www.dgn.org/leitlinien

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230219

Temporal Context Matters: An Explainable Model for Medical Resource Utilization in Chronic Kidney Disease

Omar HAMED[a,1], Amira SOLIMAN[a] and Kobra ETMINANI[a]
[a] *Center for Applied Intelligent Systems Research, Halmstad University, Sweden*

Abstract. The prediction of medical resource utilization is beneficial for effective healthcare resource planning and allocation. Previous work in resource utilization prediction can be categorized into two main classes, count-based and trajectory-based. Both of these classes have some challenges, in this work we propose a hybrid approach to overcome these challenges. Our initial results promote the value of temporal context in resource utilization prediction and highlight the importance of model explainability in understanding the main important variables.

Keywords. Resources Utilization, XAI, Electronic Health Records, Deep Learning

1. Introduction

The availability of medical resources is essential to provide a quality care service. These resources are expensive and can be in short supply. Forecasting resource utilization can help healthcare providers and governments estimate future utilization for each patient, especially in cases of chronic diseases. Previous methods for forecasting resource utilization can be grouped into count-based and trajectory-based [1,2], both using medical and administrative variables from electronic health records (EHRs). However, count-based methods ignore the temporal context and health status progression, while trajectory-based methods face challenges with irregular time spans and lengthy sequences [2]. To address these issues, this paper proposes a new hybrid approach that considers temporal context while avoiding lengthy sequences and irregular time spans.

2. Methods

This study is based on retrospective data from EHR in Region Halland, Sweden. The cohort consisted of 27,519 patients diagnosed with chronic kidney disease (CKD) according to ICD-10-SE, estimated glomerular filtration rate (eGFR) and urine albumin-creatinine ratio (uACR) lab values [1], between January 1, 2015, and December 31, 2019. The outcome is to predict only the count of next year visits to specialists based on data from the previous year. We adopted a deep learning approach using Long Short Term Memory (LSTM) architecture. Additionally, to unveil the reasoning behind the model decision, we used SHapley Additive exPlanations (SHAP) as Explainable AI (XAI) technique. The input variables consisted of medical codes (diagnoses, procedures, medications, and abnormal lab codes), CKD-related variables (CKD stage, eGFR and

[1] Corresponding Author: Omar Hamed, E-mail: omar.hamed@hh.se.

uACR lab values), and non-medical variables (such as patient age, gender, and previous year care consumption). In order to overcome very long sequences and huge variance in the length of encounters-sequence, all variables were aggregated over a fixed time window. In our experiments we investigated one, two, four, and six months aggregation periods. The same data is used in all experiments, yet the representation differs with respect to aggregation period.

3. Results

The proposed method was compared to trajectory-based approach (considering all patient encounters) and count-based approach by aggregating historical data and feeding it into a Multilayer Perceptron (MLP) neural network. All models were trained using 5-fold cross-validation. The evaluation metrics used were area under the curve of receiver operating characteristic (ROC-AUC) and the harmonic mean of precision and recall (F1-score). Table 1 shows that the proposed method outperformed the other approaches, with the best performance obtained when using a one-month aggregation period, highlighting the importance of considering more temporal context. Using a six-month aggregation period resulted in the worst performance.

Table 1. Performance metrics of different approaches.

Aggregation period	AUC	F1
Trajectory-based (sequence of visits)	71.8 ± 0.34	70.3 ± 0.59
Count-based (MLP)	73.8 ± 0.20	73.0 ± 0.43
Our method (6 months)	74.5 ± 0.11	73.1 ± 0.58
Our method (4 months)	74.6 ± 0.11	73.5 ± 0.52
Our method (2 months)	74.7 ± 0.42	74.2 ± 0.57
Our method (1 month)	**74.9 ± 0.35**	**74.3 ± 0.57**

4. Discussion and Conclusions

The same data is used in all experiments, yet the representation differs with respect to aggregation period. The initial results show that temporal context embedded in the patient's trajectory can help resource utilization prediction models to perform better by being aware of the changes of patient health state across time. Our work also highlights the importance of model's explainability in resource utilization hence using SHAP method helped to learn about the most important variables that influence model predictions. Next steps are to investigate considering a longer period of patient history and to compare the performance of different state-of-the-art deep learning architectures such as transformers.

References

[1] Levin A, Stevens PE, Bilous RW, Coresh J, Francisco ALMD, Jong PED, et al. Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International Supplements. 2013 Jan 1;3(1):1–150.

[2] Morid MA, Kawamoto K, Ault T, Dorius J, Abdelrahman S. Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230220

An Integrated Approach to Automated Diagnosis of Cervical Intraepithelial Neoplasia in Digital Histology Images

Bríd BROSNAN[a], Inna SKARGA-BANDUROVA[a,1], Tetiana BILOBORODOVA[b] and Illia SKARHA-BANDUROV[c]
[a] O*xford Brookes University, UK*
[b]*Saarland University of Applied Science HTW Saar, Saarbrücken, Germany*
[c]*University of Oxford, UK*

Abstract. The study proposes an integrated approach to automated cervical intraepithelial neoplasia (CIN) diagnosis in epithelial patches extracted from digital histology images. The model ensemble, combined CNN classifier, and highest-performing fusion approach achieved an accuracy of 94.57%. This result demonstrates significant improvement over the state-of-the-art classifiers for cervical cancer histopathology images and promises further improvement in the automated diagnosis of CIN.

Keywords. Automated diagnosis, cervical intraepithelial neoplasia (CIN), histology image, deep learning, convolutional neural network (CNN), fusion

1. Introduction

Cervical cancer is a significant global challenge recorded as the fourth most common cancer among women globally, with estimated 604,000 new cases and 342,000 deaths in 2020 [1]. Cervical intraepithelial neoplasia (CIN) precedes invasive cervical cancer and begins with minimal structural abnormality progressing through stages of more significant abnormalities to invasive squamous cell carcinoma. Advances in AI and deep learning in digital pathology have been proven to have a significant improvement in diagnostic capabilities [2]. However, many AI-based solutions are still faced with many challenges. This project aimed to design an AI system able to cope with some of them and automate the detection of the CIN grade from digital histology images.

2. Methods

For this study, we used 1715 images from two datasets, own the CHI dataset [3] and samples obtained from the MTCHI dataset [4], to augment the CHI. MTCHI samples were derived from 80 digital images of histology whole slide images (WHI).

[1] Corresponding author: Inna Skarga-Bandurova, School of Engineering, Computing and Mathematics, Oxford Brookes University, Wheatley Campus, Oxford, OX33 1HX, UK; E-mail: iskarga-bandurova@brookes.ac.uk.

The technical context for automated CIN grading included the preparatory, detection, and decision-making phases. Data imbalance was handled, and samples were verified by expert pathologists. Seven ad-hoc CNN architectures were evaluated to define the best-performing architecture. Finally, the following data fusion techniques were applied to estimate the final CIN grade: A1 – the most severe CIN grade, even if detected in only a small part of the epithelium, is the confirmed stage of the disease; A2 – the most frequent occurrence of a grade; and A3 – an extension of A2 by using PCR rule 5 to determine predicted class confidence.

3. Results and Discussion

Repeated refinement of the data, and training parameters, based on evaluation results, enabled achievement of 92.14% accuracy for the model [5], further selected as a benchmark with mean F-score (92.19%), indicating excellent agreement between the model's predictions and the true CIN values. Mean kappa (89.4%) and weighted kappa (90.83%) also indicate good agreement. These results suggest that the slight class imbalance is not a critical in this study, where CIN 2 is the majority class followed by CIN 3, Normal and CIN 1.

The cervical WSI image SSE extraction model [5] adopted for multiclass classification of histopathology images and combined with A2 fusion demonstrated the best performance to detect CIN 1 and CIN 2 in this dataset. However, the fusion method comparison was limited because only a few samples had conflicting predictions within their patch group. Individual misclassification findings include a non-neoplasm sample misclassified as CIN 3 by all fusion methods. One case was particularly extreme because the classifier predicts three different classes for its four patches, none of which are the actual class. Assessment of each CIN class shows that the CNN model best detects CIN 3. However, precision, the proportion of positive CIN 1 classes correctly identified, is 87.65%. Recall, the proportion of actual positive CIN 2 cases correctly identified, is 84.7%. Summing up the above, although overall performance is very good, the potential application of the classifier in medical diagnostics needs further investigation. Expert validation of the classifier output would provide more confidence in the model prior to fusing the patch results and assigning the fused predicted CIN grade to the parent sample. Ultimately, it is hoped that both stages of the complex histology image analysis project will be combined, enabling CIN grade identification directly from the histology sample image.

References

[1] World Health Organisation (2022) *Cervical cancer.* Available at: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer (Accessed: 6 May, 2022).
[2] Li C, Chen H, Li X, et al. A review for cervical histopathology image analysis using machine vision approaches. Artificial Intelligence Review. 2020 Oct;53:4821-62.
[3] Biloborodova T, et al. Region of Interest Identification in the Cervical Digital Histology Images. InProgress in Artificial Intelligence: 21st EPIA Conference on Artificial Intelligence, EPIA 2022, Lisbon, Portugal, August 31–September 2, 2022, Proceedings 2022 Sep 13 (pp. 133-145).
[4] Meng Z, et al. A cervical histopathology dataset for computer aided diagnosis of precancerous lesions. IEEE Transactions on Medical Imaging. 2021 Feb 18;40(6):1531-41.
[5] Sornapudi S, Addanki R, Stanley RJ, et al. Automated cervical digitized histology whole-slide image analysis toolbox. Journal of Pathology Informatics. 2021 Jan 1;12(1):26.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230221

Digital Information Management for Advanced Practice Nursing: Needs Assessment

Heidi LAINE[a,1], Asta HEIKKILÄ[b], Laura-Maria PELTONEN[c] and Jouni SMED[a]
[a] *Department of Computing, Turku, Finland*
[b] *Turku University Hospital, Turku, Finland*
[c] *Department of Nursing Science, Turku, Finland*
Asta Heikkilä https://orcid.org/0000-0003-1151-7382
Laura-Maria Peltonen https://orcid.org/0000-0001-5740-6480
Jouni Smed https://orcid.org/0000-0002-7242-5791

Abstract. The aim of this pilot study was to explore needs related to a quality dashboard for advanced practice nursing to support quality management in a Finnish university hospital.

Keywords. Hospitals, nursing quality dashboard, nursing sensitive indicators

1. Introduction

The aim of this pilot study was to explore needs related to a quality dashboard for advanced practice nursing to support developmental work and quality management. This included interviewing Advanced Practice Nurses (APNs) from one Finnish university hospital about their information needs and experiences with their current workflow related to information management. Quality dashboards provide information about key performance indicators (KPIs) at the department or organization level to help decision making [1]. Studies show that use of healthcare dashboards can reduce cognitive burden, reduce errors, improve care processes, outcomes, and situational awareness [1, 2, 3].

2. Methods and Results

We used a qualitative study using semi-structured focus group interview, because it provides insight into topics such as revealing points of agreement, conflict or uncertainty [4]. The healthcare data of the studied university hospital is stored in a data lake. APNs can obtain information from it by information requests and the results are then provided as data tables. Five people working in varying roles related to nursing quality development participated in the interview, which lasted ca. 75 minutes and was arranged

[1] Corresponding Author: Heidi Laine, Department of Computing, University of Turku, Turku, Finland, E-mail: heidi.k.laine@utu.fi.

by using a video conference software [5]. Administrative approval for this study was given by the hospital district (T147/2022).

The results can be grouped into three categories. *Providing timely access to right information.* APNs often lack an opportunity to see the original data, regardless of its origin. The interviewees reported communication issues and difficulties selecting the right parameters for information request. Also, data obtained from the information systems can be inadequate or incorrect. The lack of valid and timely information causes interruptions and makes the study participants question whether they can trust the received data. *Providing aggregated information based on user needs.* The needed information is often distributed in several information systems that do not communicate with each other. This makes combining and comparing related data difficult. The number of information systems used in one unit to find the needed information is 10-20. Also, different professionals and departments have dissimilar information needs relating to both its coarseness and real-timeness. In some units, real-time information is vital whereas for other the validity and reliability of the data needs to be ensured. *Selecting and visualizing dashboard content.* The data currently available often lacks the context, which makes it less informative. Distinguishing information from qualitative data needs to be done manually with certain keywords since there is no statistics available. To avoid information overload the most important KPIs should be prioritised in the dashboard.

3. Discussion and Conclusions

This pilot study shows a need for more accessible and validated information presented in an easy-to-interpret form. Based on the results we can assume that the development of a nursing quality dashboard could benefit nursing quality management and development in this university hospital. One should consider the information needs of the APNs and issues experienced in the current work setting when designing the dashboard. The results align well with the existing literature: data stored in several inconsistent source systems, poor data quality and reliability and selecting suitable dashboard content are common issues in designing dashboards for healthcare [1, 6]. Since this was a pilot study with a limited setting, we need to further study the different user needs and how the dashboard could be integrated into the workflow of the APNs.

References

[1] Rabiei R, Almasi S. Requirements and challenges of hospital dashboards: a systematic literature review. BMC Medical Informatics and Decision Making. 2022;22:1:1-287. Doi: 10.1186/s12911-022-02037-8

[2] Weggelaar-Jansen AM, Broekharst D, de Bruijne M. Developing a hospital-wide quality and safety dashboard: a qualitative research study. BMJ Quality & Safety. 2018;27:12:1000–1007. Doi: 10.1136/bmjqs-2018-007784

[3] Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J et al. Dashboards for improving patient care: Review of the literature. International journal of medical informatics. 2014;84:2:87–100. Doi: 10.1016/j.ijmedinf.2014.10.001

[4] Liamputtong P. Focus group methodology: principles and practice. London: SAGE; 2016. ISBN:9781473957657 (online)

[5] In preparation: Laine H. (2023). Datasta laatua: tiedonhallinnan työkalu hoitotyön kehittämiseen. MSc Thesis.

[6] Ghazisaeidi M, Safradi R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of Performance Dashboards in Healthcare Sector: Key Practical Issues. Acta Informatica Medica. 2015;23:5:317-321. Doi: 10.5455/aim.2015.23.317-321

Section 5

Human Factors and Organisational Issues

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230223

Me-to-We Design: A Blueprint for Enriching Welfare Technologies

Morten HERTZUM[a] and Mika YASUOKA[a,1]
[a] *Department of People and Technology, Roskilde University, Denmark*

Abstract. Current technologies for ambient assisted living leave underexploited that social interaction is key to human wellbeing. Me-to-we design provides a blueprint for enriching such welfare technologies with social interaction. We present the five stages of me-to-we design, illustrate how it may transform a common class of welfare technologies, and discuss the distinguishing features of me-to-we design. These features include scaffolding social interaction around an activity and supporting transitions among the five stages. In contrast, most current welfare technologies support only some of the five stages and, thereby, either bypass social interaction or presuppose that social relations already exist. Me-to-we design offers a blueprint for building social relations stage by stage if they do not exist up front. It is for future work to validate whether the blueprint in practice delivers welfare technologies that are enriched by its profoundly sociotechnical approach.

Keywords. Welfare technology, ambient assisted living, me-to-we design

1. Introduction

Welfare technologies (outside of Scandinavia also known as *ambient assisted living technologies*) are technologies for assisting elderly and frail citizens in being more autonomous in their activities of daily living [1]. Examples of welfare technologies include assistive devices such as companion robots for cognitive stimulation, smart-home technologies such as motion sensors for fall detection, and information technologies such as online meetings for consultations with general practitioners. These technologies aim to serve the double purpose of improving the citizens' quality of life and freeing up resources in home care and other health services. Yet, current welfare technologies tend to focus on single-person use and leave underexploited that social interaction is immensely important to human wellbeing. This paper proposes me-to-we design [2] as a blueprint for enriching welfare technologies with social interaction.

The connection between social interaction and wellbeing is partly direct and partly mediated through the impact of social interaction on physical health [3]. One explanation for this connection is that social interaction fosters fellow feeling – a sense of mutual identification and sympathy – and that this feeling is central to why things matter to people, thereby instilling meaning and motivation [4]. Thus, it appears that social interaction can boost the intended effect of welfare technologies in at least three ways: through its effect on wellbeing, through its effect on physical health, and through its effect on motivation. In the following, we present me-to-we design, describe how it could transform a common class of welfare technology, and discuss its distinguishing features.

[1] Corresponding Author: Mika Yasuoka, Universitetsvej 1, Bldg 10.1, Denmark, E-mail: mikaj@ruc.dk

2. Method

To create the blueprint for enriching welfare technologies, we applied the method of design thinking, in particular its ideation process [5]. The starting point in our ideation process was to expand the design object from the technology as such to the full sociotechnical system. By including considerations about the context of use in our design thinking, it became apparent that social interaction tended to be underutilized in current welfare technologies in spite of its immense importance to human wellbeing. As a result, our ideation process converged on incorporating social interaction in the design of welfare technologies and, specifically, on a blueprint for me-to-we design.

3. Results

3.1. The Blueprint: Me-to-We Design

Me-to-we design was originally devised to reconnect museums with their audiences [2]. Its basic principle is to replace single-person activities with personally rewarding, social activities. Me-to-we design posits that such activities can be fostered by exploiting that personal entry points are an effective means of scaffolding social activities. That is, a personal entry point (me) provides the groundwork for erecting progressively more social activities (we). To move from individual toward social activities, me-to-we design provides a blueprint with five stages, each presupposing the lower-level stages:

1. *Individual consumes options:* At this stage, the use situation is construed as a single-person activity where the user consumes pre-set options without much provision for influencing them. In a museum context, this stage means that visitors are provided with access to the content they individually seek.
2. *Individual interacts with options*: At this stage, the use situation is still a single-person activity, but the user is involved in constructing and reconstructing the situation. For example, the museum visitor is provided with opportunities for asking questions and taking individual action.
3. *Individual interactions are networked in aggregate*: At this stage, the outcome of a user's activities is made available to others for inspiration and motivation. For example, museum visitors can at this stage see what others have attended to and where their own interests fit with the wider visitor community.
4. *Individual interactions are networked for social use*: At this stage, the individual user connects with other people who have the same interest and undertake similar activities. For example, museum visitors connect digitally to share experiences from their individual museum visits.
5. *Individuals engage with one another socially*: At this stage, users meet physically to pursue activities related to their shared interest. In the museum context, this stage means that visitors experience the museum as full of potentially enriching social encounters.

The ideal is not that everybody should reach the fifth stage. Different users will prefer different stages. The proposition of me-to-we design is that each stage will attract some users and that most users will make use of multiple stages.

3.2. Vision for Me-to-We Transformation of Training Technology

In Denmark, training technology is the class of welfare technologies that has contributed most benefit in recent years [6]. It is used for health promotion, preventive care, rehabilitation, and the like in still more Danish municipalities. The training technologies in use include apps, sensors, virtual reality (VR), and video consultations. An example is the app Exorlive, which provides its user with ready access to a large catalog of training exercises that are explained in videos.

Training technologies such as Exorlive are at Stage 1. Their purpose is to make it possible for users to train on their own. Rather than receiving instructions and encouragement from a healthcare professional or a fellow patient, the user individually follows on-screen instructions. If an instruction is not understood the first time, the video can be re-viewed. These technologies make it possible for individual users to train when it suits them, but there is no social component to help sustain motivation or detect errors in how users perform the exercises. Training easily becomes a dreary chore even though VR training technologies often introduce a game element.

At Stage 2, the technology must be responsive to the user's performance. Exorlive is not. In contrast, some game-like training technologies have levels with increasingly demanding exercises and dynamically assign the user exercises that match their evolving abilities. By tracking the user's performance and providing statistics about it, the technology helps the users stay disciplined about their training. The technology may also provide a facility for consulting a healthcare professional with training-related questions, such as whether an exercise should be aborted if it causes pain. Training is still a single-person activity, but it has become better matched to the individual user.

To move to Stage 3, the technology should provide for sharing the individual user's statistics with those of other users. Even without connecting directly with these other users, their aggregate performance creates a social context for the individual users' perception of their own training. They will be encouraged by seeing that others make progress. Similarly, receiving information that other users do their exercises will likely strengthen the individual users' motivation to do their own exercises: If they can do it, then maybe I can too. This way, the training is embedded in a social context, but merely a rudimentary one because it lacks direct interaction with other people.

Stage 4 involves forming online training teams. The technology must provide for forming such teams and for them to meet online. Team members may, for example, arrange to meet online twice a week to do their exercises. The team members remain in their individual homes, but they become accountable to one another for attending their joint online training sessions. By having these sessions as video meetings, the team members make their training effort visible to one another and create occasions for social interaction to arise – about the training, their homes, and so forth. Training is no longer just about the individual user's personal exercise; it is also about being part of the team.

At Stage 5, team members meet with one another for face-to-face training sessions or to take part in events that mark important milestones in their rehabilitation, such as being able to go out dancing again. Technology must support the teams in scheduling their sessions, including support for forming teams with members who live in the same neighborhood. Meeting face to face creates further possibilities for social interaction and, thereby, embeds the training in a potentially rich social context that brings together people with similar health issues. Training has been transformed from something that requires personal discipline to a socially driven experience – from me to we.

4. Discussion

We acknowledge that multiple welfare technologies and research projects within ambient assisted living focus on social interaction [e.g., 7–9]. What me-to-we design offers is, partly, a conceptual framing for reflecting on how these technologies incorporate social interaction and, partly, a blueprint for deriving more benefit from social interaction in future welfare technologies. Below, we elaborate on five features of this blueprint.

First, me-to-we design *takes a sociotechnical approach* to the design of welfare technology. The design of sociotechnical systems is demanding because of the large number of variables. However, it may also be rewarding because it is through their inscription in social contexts that technologies become interlinked with the things that matter to people. We contend that current training technologies tend to be designed for single-person use (Stage 1) and have described how they could benefit from me-to-we design to avoid becoming monotonous and underused. However, we also acknowledge the presence of welfare technologies designed for connecting geographically dispersed families, combating loneliness among elderly people living alone, and calming the user through social interactions [8]. These technologies already attend to social issues (Stages 3 and 4), but they tend to focus exclusively on enabling conversation.

Second, me-to-we design *scaffolds social interaction around an activity*. Thereby, the social interaction is about something. This scaffolding is particularly important when the technology aims to create interactions among people who do not know one another beforehand [7]. It is easier to start talking with someone about a current and shared activity, such as training, than to start from scratch. However, the scaffolding around an activity may also enrich interactions with family and friends. For example, family and friends may livestream activities to include elderly people through real-time virtual presence when physical presence is not possible [10]. This way, elderly people can join family outings to loved places or attend their grandchildren's graduation ceremony. Such real-time inclusion in activities through livestreaming is more fulfilling than a post hoc narration during a video call set up with an exclusive focus on conversation.

Third, me-to-we design *invites diverse entry and end points*. The preferred balance between social interaction and focal activity will differ across users and technologies. The me-to-we model does not prescribe that the user should start at Stage 1 and end at Stage 5. Rather, welfare technologies should allow for users to start and end at the stages that best match their preferences. To do so, the technologies must provide support for all five stages. One way of achieving this goal is by supplying multiple technologies, each focusing on a subset of the five stages. Alternatively, it may be possible to devise generic components for extending a welfare technology with functionality supporting social interaction (Stages 3 to 5).

Fourth, me-to-we design *supports transitions from one stage to another*. For example, training technologies that extend individual users' access to their own statistics with possibilities for sharing them with other users support transitioning from Stage 2 to 3. Similarly, technologies that enable livestreaming support transitions back and forth between real-time inclusion in activities (Stage 4) and merely talking about the activities in the aggregate (Stage 3). This way, me-to-we design leads to technologies that support the transition from single-person use to communal use. In contrast, welfare technologies such as the telepresence robot OriHime [9] and online consultations with healthcare staff [6] presuppose that the users already have a relation and a recognized need for communicating with one another. By skipping the lower stages of the me-to-we model, these technologies do not support the gradual building of social relations.

Fifth, me-to-we design *attends to the social*. It provides a blueprint for fostering social interaction and exploiting its positive effects on motivation, health, and wellbeing. By attending to the social, me-to-we design provides needed contrast to the many technologies for surveilling elderly people in their homes for safety reasons [11], motivating them with exergames for training reasons [12], or stimulating them with social robots for companionship reasons [13]. These technologies pose ethical dilemmas about privacy and pseudo-social interaction. Rather than exploiting the positive effects of social interaction, these initiatives bypass it to preserve human resources, which are presumed to be scarce. Me-to-we design challenges this presumption by seeking to create meaningful social interactions among users who are engaged in similar activities.

5. Conclusion

It is challenging to design welfare technologies that truly assist elderly and frail citizens. We have proposed me-to-we design as a blueprint for enriching such technologies. Me-to-we design contributes ideas and stages for inscribing welfare technologies in social interaction and, thereby, interlinking them with things that matter to people‘s health, motivation, and wellbeing. To assist elderly and frail citizens in an effective and fulfilling manner, the authors would like to test the validity of the blueprint in future work.

References

[1] Blackman S, Matlo C, Bobrovitskiy C, et al. Ambient assisted living technologies for aging well: a scoping review. Journal of Intelligent Systems. 2016 Jan 1;25(1):55-69.

[2] Simon N. The participatory museum, Museum 2.0, Santa Cruz, CA, 2010.

[3] Helliwell JF, Putnam RD. The social context of well–being. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 2004 Sep 29;359(1449):1435-46.

[4] Sayer A. Why things matter to people, Cambridge University Press, Cambridge, UK, 2011.

[5] Dell'Era C, Magistretti S, Cautela C, et al. Four kinds of design thinking: From ideating to making, engaging, and criticizing. Creativity and Innovation Management. 2020 Jun;29(2):324-44.

[6] Center for Welfare Technology, Statusrapport 2020: Velfærdsteknologi i kommunerne, Local Government Denmark, Copenhagen, 2020.

[7] Báez M, et al. Online group-exercises for older adults of different physical abilities. In2016 International Conference on Collaboration Technologies and Systems (CTS) 2016 (524-533). IEEE.

[8] Thangavel G, et al. Customized information and communication technology for reducing social isolation and loneliness among older adults: scoping review. JMIR Mental Health. 2022;9(3):e34221.

[9] Vikkelsø S, et al. The telepresence avatar robot OriHime as a communication tool for adults with acquired brain injury: an ethnographic case study. Intelligent Service Robotics. 2020 Oct;13:521-37.

[10] Kim S, Junuzovic S, Inkpen K. The nomad and the couch potato: Enriching mobile shared experiences with contextual information. InProceedings of the 2014 ACM International Conference on Supporting Group Work 2014 Nov 9 (pp. 167-177)..

[11] Vermeer Y, et al. What do we require from surveillance technology? A review of the needs of people with dementia and informal caregivers. Journ of rehab and assist techn engin. 2019;6.

[12] Göbel S, Hardy S, Wendel V, et al. Serious games for health: personalized exergames. InProceedings of the 18th ACM international conference on Multimedia 2010 Oct 25 (pp. 1663-1666).

[13] Lin CC, Liao HY, Tung FW. Design guidelines of social-assisted robots for the elderly: a mixed method systematic literature review. InHCI International 2020–Late Breaking Papers: Cognition, Learning and Games: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22 2020 (pp. 90-104). Springer International Publishing.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230224

Usability Engineering of Dynamic Biosignal Displays Using Ventilation Data

Beatrice COLDEWEY[a,1], Anne Marie KLÖCKENER[a], Christof GÖBEL[b], Rainer RÖHRIG[a] and Myriam LIPPRANDT[a]
[a] *Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Germany*
[b] *Löwenstein Medical Technology GmbH + Co. KG, Hamburg, Germany*
ORCiD ID: Beatrice Coldewey: 0000-0001-9564-9467, Anne Marie Klöckener: 0000-0002-6984-4362, Rainer Röhrig: 0000-0002-0032-5118, Myriam Lipprandt: 0000-0001-9371-0551

Abstract. The aim of this work is to develop and evaluate a multi-stage procedure model for the identification of use problems and optimization of usability using biosignal data. The concept is divided into 5 steps: 1. static analysis of data to identify use problems; 2. conducting interviews within the context of use and requirements analysis to investigate problems in more detail; 3. developing new interface concepts to implement the requirements and a prototype of an interface including dynamic visualization of data; 4. formative evaluation using an unmoderated remote usability test; 5. usability test with realistic scenarios and influencing factors in the simulation room. The concept was evaluated in the ventilation setting as an example. The procedure allowed the identification of use problems in the ventilation of patients as well as the development of suitable concepts and their evaluation to counteract use problems. To relieve users, ongoing analyses of biosignals with respect to the use problem are to be carried out. To overcome technical barriers, further development is needed in this area.

Keywords. Usability Engineering; Medical Device; Data Analysis, Data Visualization; Prototyping

1. Introduction

The usability of medical devices such as vital signs monitors or ventilators is safety and time critical for monitoring and treating patients. Use errors can lead to a hazardous situation and patient harm. The design of user interfaces contributes significantly to the prevention of use errors [1,2]. Use errors in mechanical ventilators are caused by the design of interface elements, the implementation of the task navigation or the representation of medical functions, for example, the naming of ventilation modes [3]. First paragraph.

Mechanical ventilators support patients with insufficient or no self-breathing. Ventilation parameters measured, e.g., respiratory rate, tidal volume, and ventilation pressures depend on the patient's condition and the ventilator settings. The discrete time series values are obtained by sensors at constant time intervals by sampling the

[1] Corresponding Author: Beatrice Coldewey, Institute of Medical Informatics, University Hospital RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany; E-mail: bcoldewey@ukaachen.de

continuous parameters. They are displayed in real-time to the medical staff as numerical value or as a graph. Alarm messages indicate value deviations outside the set limits.

Dynamic medical data not only matter in patient care. The analysis of dynamic biosignal recordings can provide retrospective insights into the quality of the medical treatment and problems that may have occurred. As part of a root cause analysis [4], conclusions can be drawn about the role of the device and use errors to counteract interface design weaknesses. In the development process, simple concept sketches up to fully developed prototypes are used as a basis for discussion and testing [5]. Realistic displays or simulations with dynamic data are necessary to verify usability in user tests.

In the field of usability engineering, little is known about the retrospective analysis of as well as prototyping using dynamic biosignal data. The aim of this work is to develop and evaluate a multi-stage procedure model using biosignal data to identify use errors and improve usability. Data from the ventilator setting and ventilator interfaces serve as an example for the evaluation of this procedure.

2. Methods

A multi-stage procedure model (Figure 1) was developed in line with the usability engineering process with a specific focus on dynamic biosignal data. It contains the identification of use errors, requirements and the context of use, the development of new interface concepts, prototypes and the performance of usability tests.

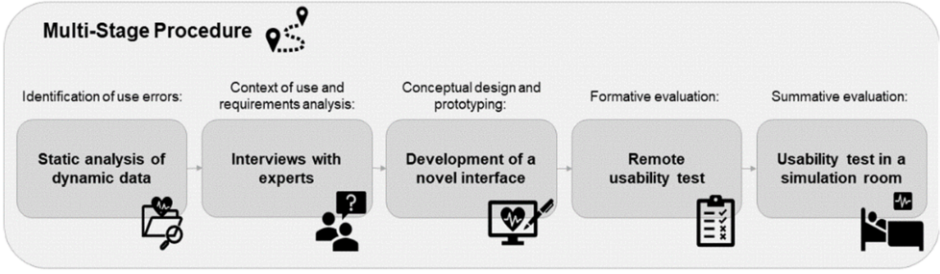


Figure 1. Model of the multi-stage procedure in line with the usability engineering process

2.1. Step 1: Static analysis of dynamic data

A detailed analysis of existing data shall reveal information regarding the quality of the medical treatment and problems that may have occurred. Biosignals available to medical staff during treatment are used only to assess the patient's condition in real-time with limited insight into historical data. A tool that allows a static view of the overall available data, annotation of conspicuous areas, and collaborative work is needed.

2.2. Step 2: Interviews with experts

In the context of the use and requirements analysis, patient-specific issues from step 1 are to be analysed in more detail with suitable methods. We decided on semi-structured interviews with device users to clarify questions regarding the investigated anomalies and medical facts. The interview results will be examined regarding new requirements on the medical device interface counteracting use errors and the context of use.

2.3. Step 3: Development of a novel interface

The requirements from step 2 are the basis for the implementation of new design solution. For the implementation of a realistic prototype of an interface, simulated data from patients are required to ensure interaction between the user and the device.

2.4. Step 4: Remote usability test

Usability evaluations are methods for testing medical devices and identifying problems during use [1, 6]. An unmoderated remote usability test was chosen for a formative evaluation. Due to the Covid-19 pandemic, but also for rapid iterative feedback, moderated or unmoderated remote tests play an increasingly important role [7]. Other formative evaluation methods to identify usability issues could also be suitable.

2.5. Step 5: Usability test in a simulation room

For a comprehensive summative evaluation, the entire socio-technical system in which the interaction with the device takes place must be considered. Influencing factors such as stressors or acoustic noise exposure can promote the occurrence of use errors. Since the application is safety-critical, the usability cannot be evaluated on real patients. In a simulated care situation, the human-machine interaction must be modelled on reality.

3. Results

The development process was evaluated using mechanical ventilation as an example. Requirements were elicited, and concepts were developed and evaluated with a focus on a ventilation problem identified in step 1. In the process, various software artefacts were used or developed. The procedure was carried out by an interdisciplinary team, further medical experts were consulted in the respective steps.

3.1. Step 1: Static analysis of dynamic data

The static visualization was implemented using the open-source software Grafana (https://grafana.com) and a MySQL database on a web server. The Grafana application provides a graphical web interface to configure the data sources and to create a dashboard containing the available data (e.g. insp. / exp. pressure, respiratory rate, inspiratory-to-expiratory ratio, tidal and minute volume, flow (Figure 2), volume). This allowed retrospective analysis, annotation and discussion of abnormalities and pathological processes in the recorded data (anonymized recordings, 2 – 8 hrs each, 50 Hz).

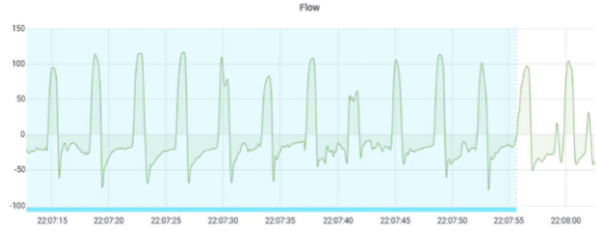


Figure 2. Excerpt of the flow graph of the annotated ventilation data in Grafana

3.2. Step 2: Interviews with experts

The interviews with seven experts (ventilator users with different professional experience, number of participants according to the saturation principle) provided detailed information for one exemplary ventilation problem regarding indicators, e.g., the shape of the curves and their detection during the treatment of patients. In addition to user characteristics, the type of information presented was seen as a factor supporting early detection or overlooking.

3.3. Step 3: Development of a novel interface

First sketches of the graphical concepts were implemented as static mock-ups. For the development of interactive prototypes, no tools were available on the market that allow a quick adaptation of the GUI and the display of dynamic data for the simulation of various scenarios. The GUI of commercially available ventilators are not modifiable from “outside”, so a GUI prototype had to be re-implemented. The patient's breathing mechanics and the ventilator mechanics, including the ventilation tube, were substituted by a bidirectional interface to a real ventilator (data provider), which is connected to a lung simulator [8].

3.4. Step 4: Remote usability test

Twelve Subjects participated and were provided with video recordings of ventilator interfaces with different ventilatory problems to be detected via the online survey tool SoSci Survey (https://www.soscisurvey.de/). The remote test allowed for initial feedback regarding the effectiveness (identification rate of ventilatory problem) and user satisfaction, as well as comments on the adapted interface design compared to the regular interface. As it was an unmoderated procedure, the efficiency could not be measured. [9]

3.5. Step 5: Usability test in a simulation room

The simulation room enables the implementation of a realistic setup including disturbing/ stressful environmental factors for the selected scenarios, e.g. a doctor's visit or the treatment of an acute patient. The evaluation of usability (effectiveness, efficiency, satisfaction and use errors) is performed over the different scenarios, each requiring interaction with different interface elements of the ventilator. A case number of 8-12 subjects allows the qualitative evaluation of a large number of possible findings [6].

4. Discussion and Conclusions

Usability engineering is a core component in the development of software systems and devices [10]. This study provides a procedure for the analysis of existing device data based on a static representation and further investigation of identified problems to improve usability of medical devices. An advantage of analysing dynamic data with static methods is that the health status and the health related progression can be better surveyed. Early pathological processes could be identified in the ventilation data for the detection of which the devices do not offer support so far. Besides the detection of

pathological conditions, it is also possible to identify human-machine interaction errors. Concrete requirements for the enhanced graphical interface could then be identified in a second step. The highlighting of pathological processes in a medical device is intended to support the user but may also contain a new source of use errors. Through the various usability tests, the effect of the newly developed interface could be evaluated with regards to use errors and hazard-related use scenarios. One hurdle is that real data is needed for our approach. In addition, access to users for interviews and participant observations is limited by the lack of time and resources among nurses and physicians.

Particularly in medicine, safe human-machine interaction is a regulatory requirement for patient safety [1, 2, 11]. For translational research and development, innovative usability approaches with biosignals are not publicly available as they are often internal to the company. Ventilators but also ECGs, EEGs or vital sign monitors, all these systems are medical devices that contain dynamic biosignals. The analysis of real ventilation data showed that existing biosignals have the potential to provide insights for research and development. To promote the evaluation of new concepts within realistic scenarios, further development of simple tools and open-source interfaces is required.

Acknowledgements

This project was funded by the German Federal Ministry of Education and Research (BMBF) (simpliVENT FKZ: 13GW0262C).

References

[1] IEC 62366-1:2015, Medical devices - Part 1: Application of usability engineering to medical devices.
[2] Food and Drug Administration. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff 2016. Washington, D.C., USA: FDA.
[3] Coldewey B, Diruf A, Röhrig R, Lipprandt M. Causes of use errors in ventilation devices - Systematic review. Applied Ergonomics. 2022;98:103544. doi:10.1016/j.apergo.2021.103544.
[4] Lipprandt M, Klausen A, Alvarez-Castillo C, Bosch S, Weyland A, Röhrig R. Erweiterte systematische Fehleranalyse zweier CIRS-AINS Alert-Fälle: Vom Anwender- zum Anwendungsfehler. Anästh Intensivmed. 2020:76–84. doi:10.19224/ai2020.076.
[5] Camburn B, Viswanathan V, Linsey J, Anderson D, Jensen D, Crawford R, et al. Design prototyping methods: state of the art in strategies, techniques, and guidelines. Design Science. 2017;3:E13. doi:10.1017/dsj.2017.10.
[6] PE ME, Kendler J, Strochlic AY. Usability testing of medical devices. CRC press; 2015 Dec 23.
[7] McLaughlin AC, DeLucia PR, Drews FA, Vaughn-Cooke M, Kumar A, Nesbitt RR, Cluff K. Evaluating Medical Devices Remotely: Current Methods and Potential Innovations. Human Factors. 2020;62:1041–60. doi:10.1177/0018720820953644.
[8] Coldewey B, Göbel C, Röhrig R, Lipprandt M. Prototyping von Beatmungsgeräte-Interfaces – Herausforderungen bei der Anzeige dynamischer Daten. German Medical Science GMS Publishing House. August 19, 2022:DocAbstr. 137. doi:10.3205/22gmds053.
[9] Coldewey B, Röhrig R, Lipprandt M. Online Usability Tool - First Experiences from Testing a Ventilator Interface. Stud Health Technol Inform. 2022 May 25:807–8.
[10] Geis T, Johner C. Usability Engineering als Erfolgsfaktor: Effizient IEC 62366- und FDA-konform dokumentieren: Beuth Verlag GmbH; 2020.
[11] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230225

Eye-Tracking on Touch Screen - Evaluating User Interaction

Camilla Panduro NIELSEN[a,1], Jacob NIELSEN [b], Chris Bang SØRENSEN [b] and Ellen Raben PEDERSEN [a]
[a] *SDU Health Informatics and Technology, University of Southern Denmark, Denmark*
[b] *SDU Game Development and Learning Technology, University of Southern Denmark, Denmark*
ORCiD ID: Camilla Panduro Nielsen 0000-0001-7547-5455, Jacob Nielsen 0000-0002-5134-8950, Chris Bang Sørensen 0000-0002-2279-2091, Ellen Raben Pedersen 0000-0001-9917-9837

Abstract. This paper suggests a setup for using remote eye-tracking on a touchscreen tablet to evaluate user interaction for older adults interacting with a user-driven hearing test. By using video recordings to support the eye-tracking data, it was possible to evaluate quantitative usability metrics that could be compared to other research findings. The video recordings revealed useful information to distinguish between reasons for gaps in data and missing data and to inform future similar studies of human-computer interaction on a touch screen. Using only portable equipment allows researchers to move to the location of the user and investigate the user interaction of devices in real-world scenarios.

Keywords. Eye-tracking, touch screen, video, usability, human-computer interaction

1. Introduction

Methods like Concurrent Think Aloud (CTA) and Retrospective Think Aloud (RTA), are often used for testing and evaluating user interaction [1]. These methods can have limitations for investigating user interaction, as CTA can cause the test subject to be distracted from the task at hand and make them perform differently, and RTA relies on the test subject to remember everything they thought while the test was happening, which is not always obtainable, especially when tasks become complex [1, 2]. Eye-tracking can be used to give quantitative measurements of user interaction to compensate for the limitations of the CTA and RTA, and it is especially useful for people with a decreased capacity in working memory, as it allows for investigation of user behavior without placing an additional cognitive burden on the participants [1].

This study aimed to set up a method for remote eye-tracking to evaluate user interaction on a User Interface (UI) of a user-driven hearing test (pure tone audiometry), displayed on a touch screen. The focus group of the study was older adults, i.e., people aged 50 years and above. Older adults fit the age of most first-time hearing aid users in

[1] Corresponding Author: Camilla Panduro Nielsen, E-mail: camilla.panduro.nielsen@rsyd.dk.

Denmark, and they are relevant for eye-tracking studies, as their working memory is lower than that of younger adults, thus making them less ideal for CTA and RTA [3, 4].

2. Methodology

To compose comparable quantitative measures for user interaction, usability metrics were set up, according to the ISO 25022:2016 standard on measurement of quality in use. The ISO standard suggests that quality in use can be measured by evaluating effectiveness, efficiency, satisfaction, freedom from risk, and context coverage [5]. In this study, effectiveness and efficiency were measured using eye-tracking and video recordings; satisfaction by assessing the user's response to the System Usability Scale questionnaire [6]; freedom from risk by observing potentially dangerous situations from the video recordings; context coverage was omitted, as this study only assessed the intended context of use. As only effectiveness and efficiency were assessed using eye-tracking, only these measures will be reported in this paper.

2.1. Experimental Setup

In this study, 17 people in the age group 51 to 85 years of age participated. None of the participants had interacted with the hearing test prior to the study. The Tobii EyeX eye-tracking controller (EyeX) was used for the experiment. It has previously been evaluated to potentially be useful for research applications [7]. The EyeX does not provide an accessible data stream of numbers, and therefore the software GazeViewer from Tobii Dynavox (GazeViewer) was used to display screen recordings with heat maps and gaze plots overlayed [8, 9]. Two GoPro cameras were used to capture the participant's movements during the hearing test. One was set behind the user to the opposite side of their dominant hand, and one was set facing the user, so their face and upper body were visible.

The hearing test used for the study was based on the Automated Method for Testing Auditory Sensitivity (AMTAS), integrated into the Interacoustics AMTAS (IA-AMTAS) solution Affinity Compact Suite. The IA-AMTAS was an alternate implementation for research purposes in the User Operated Audiometry project, with differences in UI and instructional text [10]. The UI was displayed on a 15.6-inch ZenScreen touch MB16AMT tablet, and the UI comprised several screen images that the user was interacting with. The eye-tracking controller was mounted at the bottom of the tablet, following the manufacturer's instructions [11]. The test stimuli (pure tones) were presented to the participants through RadioEar DD450 headphones.

The participants were seated at a table with the tablet at a distance so they could comfortably touch the screen, while the eye-tracker could detect the reflections from their eyes. The acceptability of the participant placement was investigated using the built-in seven-point calibration routine from the EyeX software, following the manufacturer's instructions while changing the participant's seating position, the light conditions, placement of the tablet, etc., and redoing the calibration until acceptable or abortion of the test if no acceptable calibration could be reached [9, 12].

2.2. Annotation of data

Before starting the experiments, every possible task to conduct while interacting with the UI of the hearing test was denoted with a specific code. Mandatory tasks were marked as such. All participants were faced with the same mandatory tasks and allowed to spend as much time as they wanted to interact with the UI without interruptions. Examples of mandatory tasks were: looking at information text, tapping appropriate buttons on the screen, picking up headphones, etc. After the experiments, the video recordings from the cameras and the screen recording with heat maps and gaze plots were synchronized, and annotation of task occurrence and duration was done in the transcription program ELAN by looking at the video recordings and noting which task occurred for how long [13, 14]. Annotations were done following the procedure in **Figure 1**.

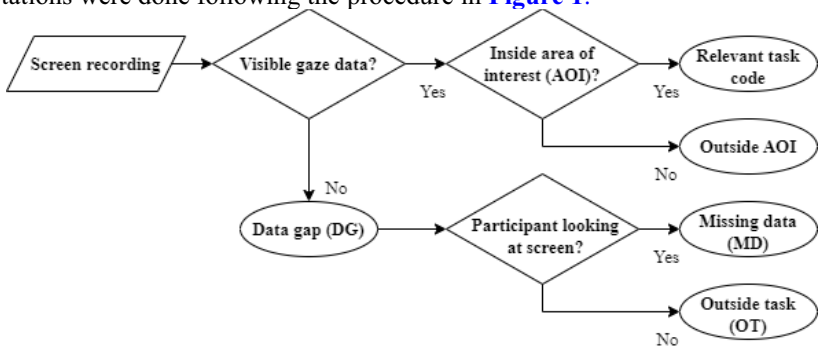


Figure 1. Flow chart describing the procedure for data annotation.

If gaze data appeared inside predefined Areas of Interest (AOI's), it was noted with the relevant task. If a mandatory task was not conducted, the duration of that task was left empty. If no gaze data were visible, it was denoted as data gaps (DG), and the DG's were further investigated by looking at the video recordings to distinguish between no data value stored from the eye tracker - missing data (MD), or the user looking outside the screen - performing outside tasks (OT's), which would not be detectable by the eye-tracker. In this study, the MD could be caused by random user behavior or prompted by the UI, so it was classified as data Missing At Random (MAR) [15]. If the recordings showed that MD occurred during the performance of a task, then all eye-tracking data were deleted for that task, using pairwise deletion [15].

The effectiveness was the percentage of mandatory tasks completed for each participant. The efficiency was reported as the percentage of time with available data when users were looking at AOI's. DG, MD, and OT were reported as percentages of the total time spent interacting with the UI for each participant.

3. Results

Table 1 shows a summary of the results. The average effectiveness is reported alongside the average percentage (%Average) of efficiency, DG, MD, and OT taking the different time spent pr. individual participant into consideration. The bottom row shows the average deviation from the average (Avgdev) for all measures. The video recordings revealed that DG, as well as OT, occurred for 16 out of the 17 participants.

Table 1. Average, percentage average, and average deviation for effectiveness, efficiency, data gaps (DG), missing data (MD), and outside tasks (OT).

	Effectiveness		Efficiency	DG	MD	OT
Average	97.6%	**%Average**	98.6%	6.8%	6.0%	0.8%
Avgdev	3.0		0.5%	7.3%	7.5%	0.6%

For six participants the DG was partly caused by MD. This is visible in **Table 1** as a higher Avgdev for both DG and MD. The reasons for MD were for three participants that their hands were blocking the eye tracker, one participant moved so that their eyes were out of range of the eye-tracker, one held the headphones so they blocked the eye-tracker, and for one participant the MD was assumed to be caused by direct sunlight.

4. Discussion

Table 1 shows that not all DG was MD, but instead, some DG was OT, accepted user behavior prompted by following the instructions on the UI. Without the cameras, all DG would have been assumed to be MD and treated as MAR, thus having to delete data for that participant. Being able to distinguish between MD and OT instead of assuming all DG to be MD, made it possible to avoid deleting relevant data while choosing the most appropriate methods for handling MD based on the true amount and cause [15].

Table 1 also shows examples of quantitative measures of user interaction like effectiveness and efficiency. Even though a data stream was not accessible as numbers, comparable measures could still be obtained by measuring from the screen- and video recordings, when following a predefined annotation procedure. It allowed for assessing some user performance metrics against other researchers' experiences. For example, the general average effectiveness of UI's has been assessed to 78% [16], and the effectiveness of the UI tested in this study was 97.6% on average. Furthermore, the efficiency shows that the UI prompted the participants to look where the researchers expected them to look 98.6% of the time on average.

Intrusive and head-mounted eye-tracking systems were excluded from the setup, as wearing the eye-trackers could make the participants more aware that they were being studied and then potentially cause them to change their behavior. The EyeX and GazeViewer alongside GoPro cameras provided measures for effectiveness and efficiency with a percentage average of 6.0% MD, which could be used for statistical analysis without affecting the bias of the results [17].

No matter which eye-tracking systems would be used, supporting the eye-tracking data with video recordings and potentially a questionnaire could reveal additional useful information about the experiments. As all the systems are portable, it allows the researcher to move to the location of the users, and this can give a more realistic experimental setup when investigating user interaction and usability on devices in real-world scenarios in any location. To determine if the usability is good/better or bad/worse, it is suggested that the researcher determines thresholds before starting the experiments.

The reasons for MD in this study were largely expected user behavior and environmental conditions that can reasonably occur in a realistic use case for the IA-AMTAS. Knowing the reasons for MD enables the researcher to evaluate if precautions should be taken in similar future setups to avoid MD, or if some of the reasons are expected user behavior and MD therefore should be accepted to keep the experimental conditions of the interaction as close to reality as possible.

5. Conclusion

This paper suggested a method for using eye-tracking alongside video recordings to evaluate quantitative usability metrics. It can provide a starting point for future studies using eye-tracking for research on user interaction on a touch screen. By supporting eye-tracking measurements with video recordings, the researcher can understand the reasons for DG and take necessary precautions to avoid them if desirable. Furthermore, the amount and cause of MD can be discovered to allow for appropriate handling of data for further statistical analyses.

Funding

This work was supported by Innovation Fund Denmark Grand Solutions 9090-00089B and the William Demant Foundation case number 19-1251.

References

[1] Bergstrom JR, Schall AJ. Eye tracking in user experience design: Elsevier; 2014.
[2] Prokop M, et al. Impact of Think-Aloud on Eye-Tracking: A Comparison of Concurrent and Retrospectiv Think-Aloud for Research on Decision-Making in the Game Environment. Sensors (Basel). 2020;20(10).
[3] Fandakova Y, Sander MC, et al. Age differences in short-term memory binding are related to working memory performance across the lifespan. Psychol Aging. 2014;29(1):140-9. doi: 10.1037/a0035347.
[4] Houmøller SS, et al. Prediction of successful hearing aid treatment in first-time and experienced hearing aid users: Using the International Outcome Inventory for Hearing Aids. International Journal of Audiology. 2022;61(2):119-29. doi: 10.1080/14992027.2021.1916632.
[5] ISO. ISO/IEC 25022:2016 Systems- and Softwareengineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use. Dansk Standard. 2016.
[6] Brooke J. SUS: A quick and dirty usability scale. Usability Eval Ind. 1995;189.
[7] Gibaldi A, Vanegas M, Bex PJ, Maiello G. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. Behav Res Methods. 2017;49(3):923-46. doi: 10.3758/s13428-016-0762-9.
[8] tobiidynavox. Gaze Viewer Tobii Dynavox [cited 2022]. Available from: https://www.tobiidynavox.com/products/gaze-viewer.
[9] Tobii. Eye Tracking Software Tobii Gaming2022 [cited 2022]. Available from: https://gaming.tobii.com/getstarted/.
[10] Sidiras C, et al. User-Operated Audiometry Project (UAud) - Introducing an Automated User-Operated System for Audiometric Testing Into Everyday Clinic Practice. Front Digit Health. 2021;3:724748.
[11] Tobii. Mounting Guide: Tobii Help; 2014 [cited 2022]. Available from: https://help.tobii.com/hc/en-us/articles/210252245-Mounting-guide.
[12] Holmqvist K. Eye tracking: a comprehensive guide to methods and measures. Paperback ed. Oxford: Oxford University Press; 2015.
[13] ELAN (Version 6.3). Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands; 2022.
[14] Sloetjes H, & Wittenburg, P. Annotation by category - ELAN and ISO DCR. Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008). 2008.
[15] Kang H. The prevention and handling of the missing data. Korean journal of anesthesiology. 2013;64(5):402-6. doi: 10.4097/kjae.2013.64.5.402.
[16] Sauro J, Lewis JR. Chapter 2 - Quantifying User Research. In: Sauro J, Lewis JR, editors. Quantifying the User Experience. Boston: Morgan Kaufmann; 2012. p. 9-18.
[17] Dong Y, Peng CY. Principled missing data methods for researchers. Springerplus. 2013;2(1):222. doi: 10.1186/2193-1801-2-222.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230226

Inter-Professional Communications During Follow-Up of Type 2 Diabetes Patients: An Exploratory Study

Romaric MARCILLY[a,b,1], Madleen LEMAITRE[a,c], Sonia MAALEM[a,] Jan BARAN[d], Jean-Baptiste BEUSCART[a] and Paul QUINDROIT[a,1]
[a] *Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France*
[b] *Inserm, CIC-IT 1403, F-59000 Lille, France*
[c] *CHU Lille, Department of Diabetology, Endocrinology, Metabolism and Nutrition Lille University Hospital, F-59000 Lille, France*
[d] *Department of General Practice, University of Lille, Lille, France*

Abstract. Follow-up of patients with type 2 diabetes mellitus (T2DM) involves several healthcare professionals. The quality of their communication is crucial for optimizing care. This exploratory work aims to characterize those communications and their problems. Interviews were performed with general practitioners (GP), patients and other professionals. Data were analyzed deductively, and results were structured through a people map. We performed 25 interviews. GP, patients, nurses, community pharmacists, medical specialists and diabetologists are the main actors of the T2DM patients' follow-up. Three communication issues were identified: difficulties in reaching the hospital diabetologist, delays in receiving reports, and difficulties for patient to transmit information. Results were discussed in terms of tools, care pathways and new roles to support communications during T2DM patients' follow-up.

Keywords. Diabetes mellitus, Communication, Ergonomics, Critical pathway

1. Introduction

In France, the prevalence of pharmacologically treated diabetes is 5% (i.e., more than 3 million subjects), with type 2 diabetes mellitus (T2DM) accounting for 90% of the patients [1]. The prevalence of T2DM increased by 2% per year over the period 2010-2015. Diabetes causes serious long-term complications, which can occur after 10 to 20 years of metabolic imbalance, including pathogenesis of diabetic complications, neuropathy, nephropathy, retinopathy, macrovascular complications, and miscellaneous complications [2].

In France, 87% of the patients with T2DM are managed by a general practitioner (GP) alone and do not consult a diabetologist [3]. Primary care health professionals are at the heart of the health care pathway. The health care pathway for patients with T2DM involves the coordinated delivery of health and social services to satisfy their prevention

[1] Corresponding Authors: Romaric Marcilly, E-mail: romaric.marcilly@univ-lille.fr; Quindroit Paul, E-mail: paul.quindroit@univ-lille.fr.

and primary care needs. This pathway involves various community, secondary, and tertiary healthcare professionals. Yet, the quality of the communication between those practitioners is an important factor for optimal care [4]. This exploratory work aims to characterize communications during the management of T2DM patients and the difficulties faced.

2. Methods

Semi-structured interviews were conducted from May to July 2022. After introducing the project and collecting participants' consent, the interviews addressed five main themes: the management of T2DM and its variability, inter-professional cooperation, communication with the hospital in the event of the patient's hospitalization, communication tools, and communication difficulties.

Participants were recruited through purposeful sampling. Purposeful sampling consists of identifying and selecting (groups of) people who have knowledge or experience of the phenomenon under study [5]. Recruitment focused initially on GPs. Then, people identified by the analysis as being involved in the process were recruited. Recruitment stopped when no new information was discovered. Interviews were audio recorded and conducted remotely or face-to-face at the convenience of the participants.

Interviews were transcribed and analyzed as they were conducted. From each interview, we extracted semantic units describing people communicating, the direction of their communications, their content and media, and their difficulties. Data were analyzed deductively from a system engineering perspective and structured through a *people map* to describe the roles involved, their interrelationships, and the media preferred [6]. Data extraction and analysis were performed by a human factors expert and cross-checked by another one. Two interviewed GPs validated the results.

3. Results

A total of 25 interviews were conducted ranging from 20 to 40 minutes: 14 GPs, 4 T2DM patients, 2 private practice nurses, 3 community pharmacists, and 2 hospital-based diabetologists.

Seven main actors were identified (Figure 1): GP, patient, medical specialist, hospital diabetologist, private practice nurse, community pharmacist, and the analysis laboratory. These actors communicate either by (e-)mail, by phone, or face-to-face during appointments, depending on the type of information to be transmitted and the context. They mainly exchange about patient condition, lifestyle, treatments, and lab results and transmit letters and reports. Temporal dependencies are not linear between actors: depending on the patients' conditions and needs of the patient, any actor can initiate communication with any other.

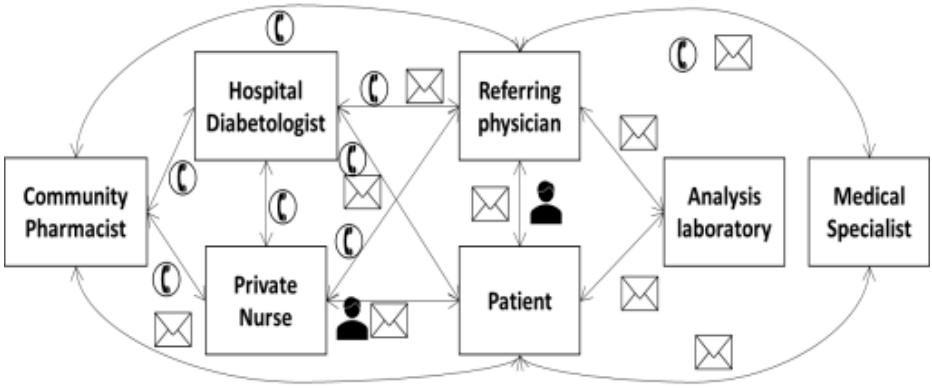


Figure 1. People map representing the communications between people involved in the management of T2DM patients. Arrows points toward the person who receives information. Icons represent the preferred mode of communication (✆, phone, ✉, mail, ♟, face-to-face)

GPs communicate with all the roles: they call the hospital when they see the patient again without having received a hospital report. They call the nurse to get information about the patient's lifestyle (activity, diet). The substitute GP may call the pharmacist for information about the patient usual treatments. The GP receives calls from hospital diabetologists who need information about the patient (history, past and current treatments, lifestyle), from nurses when a treatment no longer seems to have the desired effect, or from pharmacists in the event of doubts about a dosage or of treatment substitution. Hospital diabetologists call nurses and pharmacists to obtain information about patients' treatment or lifestyle. Pharmacists and nurses communicate on the availability of medications.

T2DM patients are a central node in the interactions between professionals. Some patients carry their medication orders with them to provide treatment information (e.g., in the event of hospitalization). Patients say that they feel responsible for passing on information between professionals and clarifying it because professionals "do not always have the information they need".

Three communication challenges were identified. First, GPs have difficulties reaching the hospital diabetologist because, for instance, of the hospital switchboard. Second, GPs do not always receive their patients' hospitalization reports on time. Sometimes, the patients inform their GP that they have been hospitalized and give them the information about the hospitalization. When GPs receive the report, they may find it incomplete (e.g., reasons for therapeutic changes). This lack of information results e.g., in prescribing tests already passed by the patient, or difficulties to optimize patient's overall treatment. Finally, T2DM patients sometimes forget to take all their orders with them to an appointment: they express difficulties to remember all their treatments and dosages accurately.

4. Discussion

Results highlight the complexity of the communications around and with the T2DM patients (Figure 1). The communication challenges identified result in increased GPs' and patient's workload, patient's responsibility for transmitting information, sub-optimal follow-up of T2DM patients (e.g., delay in treatment initiation) and rise ethical question when the GP asks the nurses information about the patient rather than directly asking the patient. The sample size in this exploratory study is small. Further studies are needed to refine the results.

Nevertheless, initial results allow us to identify several types of solutions to optimize and secure the T2DM patients' care pathway. Technological solutions (e.g., electronic patient records) should allow storing and transmitting information between professionals, and professionals and patients [7,8]. By facilitating and securing information access and transmission, these tools will support care coordination. Solutions exist [9] but technical challenges (e.g., interoperability between the information systems, data protection) [10] and ergonomic challenges (e.g., usability, acceptability) must be first overcome. Innovative care pathway solutions such as pluridisciplinary concertation meetings for drug optimization at the discharge of the patient could also facilitate the follow-up of T2DM patients: they could contribute to improve communication between GPs and hospital diabetologists and support better informed shared therapeutic decisions. New roles could also be included in the process. Advanced practice nurses (since 2018 in France) could follow patients with stabilized chronic pathologies (including diabetes) once they are discharged from hospital. Besides, they have a major role in the care coordination between healthcare professionals and patients [11].

Those instances of solutions are not mutually exclusive. They should be designed jointly in terms of roles, care pathway, and supporting information technology. The next steps will be to confirm the results and to work with representatives of patients and healthcare professionals to design together a care pathway and related roles and supporting tools with the ultimate goal of improving communication, satisfaction and patient care and safety.

Acknowledgements and Funding

The authors want to thank all the participants for their availability. This work was funded by the National Center for Precision Diabetic Medicine (PreciDIAB), which is funded jointly by the French Government (under the Investissements d'Avenir (Investments for the Future) program managed by the Agence Nationale de la Recherche (ANR, French National Research Agency; reference: ANR-18-IBHU-0001)), the European Union (through the European Regional Development Fund (ERDF; grant reference: NP0025517)), the Conseil Régional des Hauts-de-France (Hauts-de-France Regional Council; grant reference: 20001891/NP0025517), and the Métropole Européenne de Lille (MEL, European Metropolis of Lille; grant reference: 2019_ESR_11).

References

[1] Fuentes S, Mandereau-Bruno L, Regnault N, Bernillon P, Bonaldi C, Cosson E, et al. Is the type 2 diabetes epidemic plateauing in France? A nationwide population-based study. Diabetes Metab 2020;46:472–9. https://doi.org/10.1016/j.diabet.2019.12.006.
[2] Papatheodorou K, Banach M, Edmonds M, Papanas N, Papazoglou D. Complications of Diabetes. J Diabetes Res 2015;2015:189525. https://doi.org/10.1155/2015/189525.
[3] Santé Publique France. Le poids du diabète en France en 2016. Synthèse épidémiologique. Saint-Maurice, France: 2018.
[4] Ratna H. The Importance of Effective Communication in Healthcare Practice. Harv Public Health Rev 2019;23:1–6.
[5] Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 2nd ed. Los Angeles: SAGE Publications; 2011.
[6] Holden RJ, Carayon P. SEIPS 101 and seven simple SEIPS tools. BMJ Qual Saf 2021;30:901–10. https://doi.org/10.1136/bmjqs-2020-012538.
[7] Conway NT, Allardice B, Wake DJ, Cunningham SG. User Experiences of an Electronic Personal Health Record for Diabetes. J Diabetes Sci Technol 2019;13:744–50. https://doi.org/10.1177/1932296818818837.
[8] Cunningham S, Pokrajac A, Allardice B, Brillante M, Wilson L, Wake D. My Diabetes My Way: clinical outcomes impact and user experiences for an electronic personal health record for diabetes. Future Healthc J 2019;6:13. https://doi.org/10.7861/futurehosp.6-1-s13.
[9] Merlière Y. [The shared medical record, a digital health record for all]. Soins. 2020;65(842):29–32.
[10] American Diabetes Association. Summary of Proceedings of the American Diabetes Association Summit : "Overcoming Therapeutic Inertia: Accelerating Diabetes Care FOR_LIFE." Arlington, VA: 2019.
[11] Heale R, Wenghofer E, James S, Garceau M-L. Quality of Care for Patients With Diabetes and Mulitmorbidity Registered at Nurse Practitioner-Led Clinics. Can J Nurs Res Rev Can Rech En Sci Infirm 2018;50:20–7. https://doi.org/10.1177/0844562117744137.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230227

How to Design Successful Participatory Design Workshops for Digital Health Solutions?

Kerstin DENECKE[a,1] , François von KAENEL[a], Marko MILETIC[a], Carlos FERNÁNDEZ-LLATAS[b], Gema IBAÑEZ-SÁNCHEZ[b], Zoe VALERO-RAMÓN[b], Antonio MARTÎNEZ-MILLANA[b], María SEGURA[b] and Octavio RIVERA ROMERO[c,d,2]

[a] *Bern University of Applied Sciences, Medical Informatics, Bern, Switzerland*
[b] *SABIEN-ITACA, Universitat Politècnica de València, València, Spain*
[c] *Instituto de Ingeniería Informática (I3US), Universidad de Sevilla, Sevilla, Spain*
[d] *Electronic Technology Department, Universidad de Sevilla, Sevilla, Spain*

Abstract. Participatory design (PD) is increasingly used to support design and development of digital health solutions. The involves representatives of future user groups and experts to collect their needs and preferences and ensure easy to use and useful solutions. However, reflections and experiences with PD in designing digital health solutions are rarely reported. The objective of this paper is to collect those experiences including lessons learnt and moderator experiences, and to identify challenges. For this purpose, we conducted a multiple case study to explore the skill development process required to successfully design a solution in the three cases. From the results, we derived good practice guidelines to support designing successful PD workshops. They include adapting the workshop activities and material to the vulnerable participant group and considering their environment and previous experiences, planning sufficient time for preparation and supporting the activities with appropriate material. We conclude that PD workshop results are perceived as useful for designing digital health solutions, but careful design is very relevant.

Keywords. Participatory design, user-centered design, experiences, digital health intervention, user involvement

1. Introduction

Digital health interventions are increasingly developed using participatory design (PD) or user-centered design. These iterative design approaches place the individual in the center of the design; they differ in their central focus. User-centered design tries to understand the user, focusing on their needs and limitations. PD is "a process that includes the stakeholders in the early stages of design" [1], which increases their sense of ownership. PD is known to be effective in involving vulnerable user groups into the

[1] Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Institute for Medical Informatics, Quellgasse 21,CH- 2502 Biel, E-mail: kerstin.denecke@bfh.ch.
[2] Received funding from the Universidad de Sevilla and the Ministerio de Universidades of the Spanish Government under the call "Recualificación del Sistema Español de Universidades"

design process in a creative and reflective manner. Active participation allows them to concretely express their needs. In health informatics, this design approach becomes important since it is assumed to increase acceptance and usefulness of digital health solutions by end users. PD lends itself to user-centered design innovation and fits within the research field participatory health informatics (PHI) that provides resources and delivers tools supporting active participation, and focuses on individual-centered care, individual-centered self- management, and individual-centered decision making. PHI also assesses accessibility, usability, individuals' technology acceptance, experience, and satisfaction, tool appropriateness and quality [2]. While PD is becoming state of the art in developing digital health solutions, reflections and experiences with PD are rarely reported. The objective of this paper is to collect those experiences from three cases including lessons learnt and moderator experiences, and to identify challenges.

2. Methods

We conducted a multiple case study to explore the skill development process required to successfully design a solution involved in the three cases. Specifically, we aimed at obtaining a rich understanding of PD experiences. We considered three cases -of PD where we ask the moderators of the PD workshops to answer questions related to their experiences. The questions included aspects such as: challenges occurred, participant engagement, usefulness of the PD workshop results for the design / development of the digital health solution, time management, threats of the approach, and potentials for improvement. We selected three cases from health informatics in which the common point are the communication challenges with the participants that have to be handled in the design phase.

3. Experiences from three cases

3.1. PD workshop with people with intellectual disability

Several PD sessions involving people with mild to moderate intellectual disability were conducted in Valencia (Spain) within the MOVE-IT project aiming at exploring participants' perceptions, preferences, attitudes on the use of digital health solutions for physical activity promotion. Two psychologists, a physical activity expert, an occupational therapist, a physiotherapist, a logopedist, and two experts on PD designed the workshop and adapted activities to individual's skills and abilities. Several digital exergames prototypes were used as study cases to discuss usability, accessibility, personalization, gamification, and behavioral change techniques. Several canvas and accessible topic cards were designed to facilitate the discussions. 4 sessions of 70 minutes average duration were conducted in 2 occupational centers in Valencia. Three people with intellectual disability participated in each session (N=12) guided by a moderator. 2-3 members of the occupational center staff supported the moderators.

Participants appreciated to participate in the workshop and to give their opinions and preferences about a digital solution. The moderators paid special attention in trying to connect with participants from the beginning of the session to let the participants feel comfortable (e.g. by structuring the sessions similarly to others that were done regularly at the centers). Participants were engaged with the activities and liked to discuss

technological topics with the moderators. The support of the occupational center staff was extremely important to overcome communication challenges. The workshop results helped identifying relevant functionalities of the designed exergames.

The main obstacle was communicating with people with intellectual disability. Several abstract concepts such those related to behavioral change were discussed but were complex to understand by participants; providing examples of straightforward applications helped. It was sometimes unclear whether participants appropriately understood the abstract concepts or just said "yes", and it was challenging to promote deep discussion of a topic. Participants often focused on the specific characteristics of the given examples instead of the concepts. Due to cognitive fatigue, several participants easily lost their attention and it was difficult to keep them involved in the activities.

3.2. Data driven workshop

Process mining is a set of techniques to discover, analyze, and monitor a process (clinical processes, logistic analysis, patients' behaviors, etc.). Existing data from health information systems are analyzed and aggregated in dynamic visualizations. Using these technologies participatory workshops are conducted, called Interactive Process Mining Data Rodeos lasting from 30 minutes to 2 hours [3]. Experts in process mining (Process Miners), information technology staff of the institution (IT operators that are familiar with the data available in the Hospital Information System), and healthcare professionals or medical service managers are involved. The main objective is to co-create Interactive Process Indicators (IPIs) that will support domain experts in understanding the studied processes. During these sessions, the Process Miners coordinate an in vivo analysis of the processes using process mining techniques for understanding how the process is carried out, discovering actions performed, and their behaviors [4]. These techniques allow the Process Miners to select the best views representing the process (based on data provided by IT operators) that are understandable for domain experts.

To facilitate successful communication in these sessions, Process Miners must rapidly become familiar with the domain experts` terminology. They must also be familiar with the most important aspects of the study case (the process, the pathway(s), the key performance indicators, etc.). Keeping participants' interest and engagement highly depend on the Process Miners` skills. The first sessions are intended to make the domain experts aware of the process mining methodologies, and what they can expect from data rodeos. The moderator's task is to identify a general process that offers a rough overview of the process. This process will then be developed throughout subsequent sessions, leading to the co-design of a production-level indicator that might be used to analyze the process in actual daily practice. Co-design is essential for ensuring that domain experts truly understand and have trust in the final IPI. It is best to avoid delving too deeply into data curation in the initial stage. The IPIs' impact on the data quality allows for identifying data quality problems that may be caused by information system errors or process inaccuracies. This enables IT professionals to update data or address flaws in the hospital's health information systems. It also enables domain experts to understand the quality of data and how it influences the final indications.

3.3. PD workshop with persons at risk for suicide

As part of the SERO suicide prevention program a mobile app is developed to improve the self-management of suicidal individuals [5]. The concept of the app was developed

together with participant groups consisting of 8-10 participants (health professionals, persons at risk and their relatives) who provided requirements engineering and gave feedback. Methods used included brainstorming, mockup software (Figma), prototypes, usability tests. There was one iteration per functionality of the app that was provided as a minimal viable product for testing, evaluation, and feedback. In the three PD workshops, participants were split into two working groups with one moderator per group; the workshops were held online with a duration of 1 hour: 10 minutes common introduction; 40 minutes group work; 10 minutes presentation of the group works. The purpose of the workshops were collecting requirements, expectations, soft factors (colors, overall appearance, message to be carried, ...). The interaction between the participants was supported using Miro boards.

Participants had difficulties to become activated, in formulating ideas and to prioritize functionalities. The individual views on some of the project's objectives were completely divergent among participants. For the developers it was difficult to weigh the opinions of the participant groups engaged (e.g. range the input from professionals higher than those of affected persons).

The participants were very interested and found the workshops productive; it was the first time they participated in such a workshop. They extremely appreciated that they could contribute to the development of the app. The results from the workshop directly influenced the functionality and the design of the app. Design decisions were based on facts from a group of participants and not on mere opinions of individuals who had decision-making power. Threats of the approach are that the participants might not be diverse enough. To avoid influencing the participants with moderator opinions, the moderator only asked questions and dug into the details without taking a position. Care was taken to ensure that the questions asked could not be answered merely with yes or no, so that the people concerned could describe their thought processes. In addition, the underlying motivations of statements made were collected. The efficiency of the process could be improved, in particular the coordination of all involved parties.

4. Recommendations for successful PD workshops

The three cases demonstrated that PD workshops are very useful for design and development of digital health solutions, even when vulnerable groups are addressed. Given the nature of PD workshops, multiple different ideas can be collected. This leads to the problem that developers have to decide which input to consider since the input might be conflicting, or the budget is restricted and selections have to be made. For projects where applications are developed, it is possible to orientate on the specified project goals. Participants sometimes have unrealistic expectations. It is important to meet these expectations realistically and yet not to curb their enthusiasm. Good information, also about the course of the project and the results, is essential. Opposing opinions and statements from participants must be absorbed and processed. We derive the following recommendations for the workshop design: 1) Prepare for the terminology used by the participants when discussing technical concepts, 2) Adapt workshop to participants' skills and abilities: Guide discussions using carefully predefined topics, 3) Follow an iterative process (data rodeos), 4) Allocate sufficient time for preparation of the workshops.

So far, there is no comprehensive assessment available that studies the effect of joining PD workshops from a participant perspective. Our results show that participants highly appreciate the involvement. A careful preparation of PD workshops is essential

and very time consuming (roughly estimated approximately 5 times longer than the effective participatory phase). It is important to consider the peculiarities of the involved participants and create an environment that makes them feel comfortable. This can be achieved by building upon their experiences with other workshops or conducting the workshops in a well-known environment. Projects have problems in identifying a sufficient number of participants for PD workshops and even more complex, to find a diverse set of participants. An existing pool of the individuals willing to participate in PD workshops would make recruitment easier. A user group available throughout the entire project would be useful. Involving health professionals in PD workshops has to consider their time restrictions and flexibility is needed when other duties of the health professionals are ranked higher than participating in PD workshops. To ensure a positive "participant experience", we recommend giving them the opportunity to express their opinion, to simplify PD activities and adapt them to their abilities and skills and finally, create new technological solutions that can be relevant for them.
Engagement can be supported using: 1) Use of materials to facilitate the discussion (e.g., canvases, cards), 2) Use of study cases to contextualize the discussions (examples of exergames, features implemented in other solutions, etc.), 3) Use of interactive data visualization techniques to provide relevant information (provision and validation). A careful selection of supportive tools is essential to avoid accessibility and usability issues. When using IT, the participant`s digital literacy has to be taken into account.

5. Conclusions

Considering input from future users and experts in the field has potential to deliver solutions that are well adopted by the users and can support the healthcare treatment. The experiences from three case studies demonstrated this and there might be more successful examples. Given the feedback from the involved user groups, it might be interesting to assess, which effect the involvement has on their adherence and adoption. Do they change their behavior already as a reaction to the participation or do they adhere more to the developed solution? These and similar questions might be addressed in future. Another future research question could concern cultural differences and their impact on the success or challenges of user involvement and the impact of hierarchies among the participants (e.g., physicians of different hierarchy levels and nurses participating in a workshop).

References

[1] Hartson R, Pyla PS. The UX book: Agile UX design for a quality user experience. Morgan Kaufmann, 2018.
[2] Denecke K, Rivera Romero O, Petersen C et al. Defining and Scoping Participatory Health Informatics - An eDelphi Study. Methods in Information in Medicine, 2023 (in press).
[3] Fernandez-Llatas C (ed.). Interactive process mining in healthcare. Cham: Springer, 2021.
[4] van der Aalst W. Data science in action. *Process mining*. Springer, Berlin, Heidelberg, 2016. 3-23.
[5] Meier L, Gurtner C, Nuessli S, Miletic M, Bürkle T, Durrer M. SERO - A New Mobile App for Suicide Prevention. Stud Health Technol Inform. 2022 May 16;292:3-8. doi: 10.3233/SHTI220310. PMID: 35575841.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230228

Enhancing Inclusive mHealth Design for People Living with Dementia: Examples from Literature

Thomas ENGELSMA[a,b,c,1], Monique JASPERS [a,b,c] and Linda PEUTE [a,b,c]
[a] *Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Meibergdreef 9, Amsterdam, The Netherlands*
[b] *Amsterdam Public Health, Aging & Later Life, Amsterdam, The Netherlands*
[c] *Center for Human Factors Engineering of Health Information Technology, Amsterdam, The Netherlands*
ORCiD ID: Thomas Engelsma https://orcid.org/0000-0001-8889-1669
Monique Jaspers https://orcid.org/0000-0001-8040-9153
Linda Peute https://orcid.org/0000-0002-9859-0912

Abstract. The availability of mHealth technologies for older adults living with dementia is increasing. However, due to highly complex and varying clinical presentations of dementia, these technologies do not always meet their needs, wishes and capabilities. An exploratory literature review was performed to identify studies that applied evidence-based design principles or provide design choices that aim to improve mHealth design. These were categorized as a unique design choice to tackle barriers to mHealth use related to cognition, perception, physical ability, frame of mind, or speech- and language. Through thematic analysis, themes of design choices were summarized per category in the MOLDEM-US framework. Thirty-six studies were included for data extraction, leading to seventeen categories of design choices. This study pushes the need to further investigate and refine inclusive mHealth design solutions for populations with highly complex symptoms, such as those living with dementia.

Keywords. Dementia, mHealth, Human-centered design, Usability

1. Introduction

Dementia can introduce complex barriers to mHealth use related to decreased cognition, perception, physical ability, frame of mind, and speech and language [1]. The clinical presentation of dementia varies on an individual level, caused by differences in type(s) and severities of symptoms. This introduces additional issues when designing innovative digital tools to support activities of daily living and care processes for this population. Due to the variety in clinical presentation of dementia, a one-design-fits-all or universal approach will not suffice in realizing usable mHealth. This may lead to challenges for inclusive design and applying methodologies such as co-design and participatory design

[1] Corresponding Author: Thomas Engelsma, Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Meibergdreef 9, 1105AZ Amsterdam, the Netherlands; E-mail: t.engelsma@amsterdamumc.nl

[2]. It is impossible to include sufficient variations in clinical presentations when applying these methodologies. To support the use of these methodologies and enhance inclusive design, guidance should be provided to software developers and researchers during mHealth development for those living with dementia. To contribute to this guidance, the aim of this explorative literature review is to identify design choices from studies that designed, evaluated, or reviewed mHealth applications and the user-experience of users living with dementia.

2. Methods

An explorative review (through PubMed, Google Scholar, and IEEE) was performed in January 2022 to identify design choices made during the development and evaluation of mHealth tools for people living with dementia, using three search topics: mHealth, dementia and design (Table 1). Articles were eligible if the study described the design process of a mHealth application or performed usability testing and provides suggestions to improve usability. Papers were excluded if the focus was (1) not on mHealth, (2) on the prevention or diagnosis of dementia, (3) on (informal) caregivers only, or (4) on mHealth functionalities rather than design. For each eligible article, design choices were extracted by two researchers. Using the MOLDEM-US framework, that captures barriers to mHealth use for older adults living with dementia, these were grouped individually by two researchers as a design choice that tackles either cognition, physical ability, perception, frame of mind or speech- and language barriers [1]. Afterwards, consensus on this grouping was reached through discussion with the research team. Through thematic analysis, design choices were summarized per category from the MOLDEM-US framework, leading to different themes of design choices per category.

Table 1. Overview of Medical Subject Headings and keywords used per topic (mHealth, dementia, and design).

mHealth	Dementia	Design
smartphone [MeSH]	Neurocognitive Disorders [MeSH]	design*
cell phone use [MeSH]	Alzheimer Disease [MeSH]	criteria*
mobile applications [MeSH]	Dementia [MeSH]	guideline*
telemedicine [MeSH]	Frontotemporal Dementia [MeSH]	rule*
telehealth [MeSH]	Lewy Body Disease [MeSH]	usability
m-health OR mHealth	Cognitive Dysfunction [MeSH]	framework
eHealth OR e-Health	Cognition Disorders [MeSH]	recommendation
	Memory Disorders [MeSH]	requirement*
	Dementia	barrier*
	ADRD	specification*
	Alzheimer*	principle*

3. Results

After scanning the title and abstract of 733 unique citations, 36 studies were eligible for data extraction. An overview of the included studies can be found via: https://figshare.com/s/64ccd86abf7d0e09b4fd. From these studies seventeen themes of design choices emerged through thematic analysis, which are shown in bold in the sections 3.1 – 3.4.

3.1. Cognition

To tackle cognitive barriers to mHealth use, the first theme of cognitive design choices that emerged from various studies is the implementation of **action progress monitoring incorporating memory aids**. This can be applied through configurable reminders, auto-prompt features, checklists, auto-save functionalities and confirmations of successful task completions. This progress monitoring can support the user's declining working memory, recognition skills, the (diminished) ability to organize thoughts and actions, (diminished) attention and thinking speed. Minimization of steps for data entry should be considered during progress monitoring. Second theme includes the implementation of **tutorials** with short instructions. This can be of use when for example working memory and learnability skills are declining. Limited information should be shown on a single screen, providing and repeating simple step-by-step instructions. It can also be useful to filter irrelevant information if a user does not want or is unable to use certain functionalities or absorb content. Third, **personalization** is important when designing for people living with dementia, to allow adaptation of functionalities and task difficulty to a user's cognitive skills. Personalization can also be achieved by allowing the user to determine their individual information needs, by providing simple information items with the option to get more information and minimize distraction. Fourth, **easy and consistent navigation** should be ensured. This can be achieved by using linear navigation, but also by avoiding strong hierarchical menu structures. In line with personalization, more control methods may be implemented to allow the user to select their preferred method of interaction, such as drag-and-drop or tap. The last emerging theme concerns the **use of icons**. These should be consistent (same button for one functionality), representable and understandable (buttons should look like buttons). Icons can also be used to show successful task completion. When using icons, these should all be visible on one screen to prevent the need for scrolling.

3.2. Perception

With respect to perception, the first theme that emerged relates to **compartmentalization** of the information content and differentiation of objects from other visual items. This can be achieved through the use of bold colors, mixture of both typo- and iconography, and headings. Another important consideration in the design is how to provide **interactive system feedback**. Due to the complexity and variety in clinical presentation of dementia, studies have shown the implementation of text-to-speech modules, audio-based cues, and vibrations in case of decreasing visual acuity and text-based instructions and feedback in case of decreasing auditory acuity. With respect to **color use,** considerable barriers mHealth use for people living with dementia are glare, (diminished) color vision and contrast detection. Studies implemented the use of clear, color-neutral and distinguishable colors, increased contrasting colors compared to the interface background, and recommend avoiding excessively glaring colors. Due to the loss of touch sensation, **visual intuitive cues that show click-sensitive areas** should be provided, in combination with audible system feedback. Finally, to allow **processable elements,** the possibility to (automatically) magnify these elements, use large font sizes in general and "touch interface screen readers" should be considered. Moreover, when using reminders, the pop-up size should be increased when the user interacts with the mHealth app.

3.3. Frame of Mind

A frequently mentioned theme to tackle frame of mind barriers is to implement opportunities to receive **easy access to help,** through for example a help-desk or mechanisms that support recovery from errors smoothly. Second, the aspect of **time** is approached from various angles in relation to mHealth design. One angle suggests that users should have orientation to time (clock-time). Another angle state that, with respect to task completion, users should have ample time to respond or react to an interactive system, with triggers implemented for time-based tasks. Moreover, if timers are used, these should run up instead of down. Third, to ensure **positive system feedback,** it has been suggested to provide failure-free content only. Positive feedback from a system can be achieved through brief encouragements while completing tasks, the use of rewards when a user has attained a pre-set goal, and by confirming correctly taken steps. Fourth, as with cognition, **personalization** emerged as a theme to tackle frame of mind barriers. Personalization could be considered in terms of varying difficulty levels of functionalities, individual features based on needs and wishes, sets of functionalities based on capabilities, and when appropriate the privacy settings. Finally, to prevent **stigmatization**, the language used for the content in a mHealth app should be appropriate to the user characteristics.

3.4. Speech- and language

Two themes emerged related to speech- and language disabilities in relation to design of mHealth for people living with dementia. First, the **language use** should be explicit and consistent with everyday words and foreign language and technical terms should be avoided. Difficult terminology that cannot be avoided should be explained through for example a glossary. Second, related to **user-input**, the mHealth technology should adhere to recommendations for clear speech and it should have both text-to-speech and speech-to-text implemented. Finally, free text should be avoided as user-input when possible.

4. Discussion

This paper provides an overview of literature-based design choices for mHealth aimed at those living with dementia. Seventeen themes of design choices emerged, mainly related to cognition, perception, and frame of mind issues. Because the themes have emerged through a literature-based approach, it implies that the themes of design choices are considered to lead to improved user-experience of mHealth applications for the intended end-users. However, these still need further validation and refinement to further support software developers and academics in mHealth development.

Through this approach, there were no design choices identified in relation to the physical ability of end-users. This may be due to the fact that this review looked at software design rather than hardware design of mHealth. Other research indeed found that hardware choices are important for the user experience for dementia patients [3]. They concluded that electronic devices should not be heavy or slippery because of physical limitations of those living with dementia in a care setting.

Two strengths of this study were the use of multiple search engines to explore literature and the use of a framework to individual group the extracted design choices

and gaining consensus with the research team afterwards. Even though this search has been performed in January 2022, the results show a comprehensive set of seventeen design themes that may increase the usability of mHealth technologies for people living with dementia.

In future research, validation of the emerged themes will comprise heuristic evaluations of novel and currently available mHealth implementations. The results of these evaluations, violations of the proposed themes, will be compared with usability problems identified from usability testing with the actual end-users, those living with mild cognitive impairment or early stages of dementia. Also, the researchers will further explore the effectiveness, and investigate potential improvements, of usability testing methods when including people living with dementia. Usability testing methods such as the think aloud methodology may need adaption to increase the reliability of data when including people living with dementia. Completing a task and thinking aloud simultaneously requires a higher cognitive load, which might be too high for people living with dementia [4]. Moreover, previous research indicated observations and recording task completion rates and times produce the most reliable results when conducting usability tests with people living with dementia [4]. Finally, the applicability of the design choices may depend on the distinguished types of mHealth to support patients, which include apps to check personal health records, personal care apps, social networking apps, educational health apps, and apps to contact healthcare professionals [5]. The usefulness of design themes per type of mHealth should be further researched to improve applicability of this study's findings.

5. Conclusion

Seventeen themes of design choices emerged from literature that can support software developers and academics in designing mHealth for people living with dementia. By applying these themes methodologies such as co-creation and participatory design can be supported in the mHealth design process. In future work, the design themes should be refined and validated through user testing and mHealth implementation studies.

References

[1] Engelsma T, Jaspers MWM, Peute LW. Considerate mHealth design for older adults with Alzheimer's disease and related dementias (ADRD): A scoping review on usability barriers and design suggestions. Int J Med Inform. 2021 Aug;152:104494. doi: 10.1016/j.ijmedinf.2021.104494.

[2] Engelsma T, Yurt A, Dröes RM, Jaspers MWM, Peute LW. Expert appraisal and prioritization of barriers to mHealth use for older adults living with Alzheimer's disease and related Dementias: A Delphi study. Int J Med Inform. 2022 Oct;166:104845. doi: 10.1016/j.ijmedinf.2022.104845.

[3] Hung L, Chow B, Shadarevian J, O'Neill R, Berndt A, Wallsworth C, Horne N, Gregorio M, Mann J, Son C, Chaudhury H. Using touchscreen tablets to support social connections and reduce responsive behaviours among people with dementia in care settings: A scoping review. Dementia (London). 2021 Apr;20(3):1124-1143. doi: 10.1177/1471301220922745

[4] Gibson A, McCauley C, Mulvenna M, Ryan A, Laird L, Curran K, Bunting B, Ferry F, Bond R. 2016. Assessing usability testing for people living with dementia. In Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniques (REHAB '16). Association for Computing Machinery, New York, NY, USA, 25–31. doi: 10.1145/3051488.3051492

[5] Pires IM, Marques G, Garcia NM, Flórez-Revuelta F, Ponciano V, Oniani S. A Research on the Classification and Applicability of the Mobile Health Applications. J Pers Med. 2020 Feb 27;10(1):11. doi: 10.3390/jpm10010011.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230229

Using Real-Time Interaction Analysis to Explore Human-Robot Interaction

Jiye KANG[a], Su Jin KIM[a], Su Hyun MOON[a], Su Min KIM[a], Yujin SEO[a], Won Chul CHA, MD, PhD[a, b] and Meong Hi SON, MD [a, b, 1]
[a] *Smart Health Lab, Department of Digital Health, Samsung Advanced Institute for Health Science & Technology (SAIHST), Sungkyunkwan University, Seoul, Republic of Korea*
[b] *Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea*
ORCiD ID: Jiye Kang https://orcid.org/0000-0002-3725-0017

Abstract. Despite the increasing presence of social robots (SRs) in Human-Robot Interaction, there are few studies that quantify these interactions and explore children's attitudes by analyzing real-time data as they communicate with SRs. Therefore, we attempted to explore the interaction between pediatric patients and SRs by analyzing the interaction log collected from real-time. This study is a retrospective analysis of data collected in a prospective study conducted on 10 pediatric cancer patients at tertiary hospitals in Korea. Using the Wizard of Oz method, we collected the interaction log during the interaction between pediatric cancer patients and the robot. Out of the collected data, 955 sentences from the robot and 332 sentences from the children were available for analysis, except for the logs that were missing due to environmental errors. we analyzed the delay time from saving the interaction log and the sentence similarity of the interaction log. The interaction log delay time between robot and child was 5.01 seconds. And the child's delay time averaged 7.2 seconds, which was longer than the robot's delay time of 4.29 seconds. Additionally, as a result of analyzing the sentence similarity of the interaction log, the robot (97.2%) was higher than the children (46.2%). The results of the sentiment analysis of the patient's attitude toward the robot were 73% neutral, 13.59% positive, and 12.42% negative. The observational evaluations of pediatric psychological experts identified curiosity (n=7, 70.0%), activity (n=5, 50.0%), passivity (n=5, 50.0%), sympathy (n=7, 70.0%), concentration (n=6, 60.0%), high interest (n=5, 50.0%), positive attitude (n=9, 90.0%), and low interaction initiative (n=6, 60.0%). This study made it possible to explore the feasibility of interaction with SRs and to confirm differences in attitudes toward robots according to child characteristics. To increase the feasibility of human-robot interaction, measures such as improving the completeness of log records by enhancing the network environment are required.

Keywords. Pediatrics, real-time data, human-robot interaction, dialog data analysis, sentiment analysis

[1] Corresponding Author: Meong Hi Son, Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea, E-mail: meonghi.son@samsung.com

1. Introduction

Interest in the use of Social Robots (SRs) in the field of Human-Robot Interaction (HRI) is increasing. SRs are used in a variety of ways, from educational tools to patient treatment as assistant robots.[1] There are research results on the usefulness of SRs, such as increasing the communication effect with patients through SRs in hospitals. In previous studies, most of the research focused on the usability evaluation and experience of SRs, but few studies analyzed real data based on actual interaction logs.

HRI needs to be carefully evaluated. For this interaction, robots must consider not only functions but also social aspects and emotional factors.[2] For robots to be accepted by people, human-like characteristics must be inherent, and as humans form emotional bonds with those they interact with, HRI could be effective. This is why we need interaction analysis in HRI. From this study, we confirmed the feasibility of Human-Robot interactions through quantitative exploration of interactions and attitudes towards the robot.

2. Methods

This study is a retrospective study using log data that collected through SR and recorded video to investigate the interaction between SR and pediatric patients. This study was approved by the Institutional Review Board of Samsung Medical Center (SMC), a large tertiary hospital in Seoul in Korea (2021-05-085) and is listed on Clinicaltrials.gov as NCT04993599.

2.1. Participants

We enrolled voluntary participants through a tertiary hospital community for patients using a mobile channel. We recruited pediatric cancer patients aged 5–12 years old who had visited SMC between September 2021 and January 2022.

2.2. Study Design

2.2.1. Study Preparation

In this study, the Wizard of Oz method, which is a research method in HRI where a human operator simulates a computer's response to test a user's interactions with a system, was used for the interaction between the robot and the child. The robot used in this study was named Liku and it was developed by Torooc Inc. (Seoul, Korea). Additionally, we developed an application that can be used to apply the Wizard of Oz method, which allowed us to control Liku's various expressions and actions to communicate with children.

2.2.2. Data collection

This study used two datasets: One is interaction log data collected during the interaction between the SR and the child. The other is the video data to record the process of study during the interaction. The first dataset stored conversation in real-time as text logs in a

database. For the second dataset, the researcher converted the video data into text and stored it. Then, the two interaction log datasets were compared and analyzed.

2.2.3. Outcome Measurements and Analysis

The primary objective of this study is to confirm the level of interactivity in the interaction between the SR and the child through the analysis of interaction log data. The secondary objective is to assess the child's attitudes towards the SR through sentiment analysis of the interactions.

2.2.4. Conversation analysis

We measured the number of interactions between the SR and the child and checked the number of missing logs in the database. We also measured the time delay that occurs when saving the logs on the server. Additionally, we analyzed the similarity of sentences using the Terminology Frequency-Inverse Document Frequency (TF-IDF) method to check the similarity between the log data in the server and in the video.

2.2.5. Sentiment analysis

We utilized an API based on commonly known sentiment analysis models to detect positive, neutral, and negative sentiments from the collected interactions. Through this analysis, we aimed to investigate how the sentences and words used in HRI affect the sentiments expressed. Additionally, the characteristics of the child were observed by pediatric psychologists, and four main types were identified: ‘first response to the robot’, ‘general response and attitude’, ‘basic emotion’, and ‘interaction initiative’.

3. Results

3.1. Participants’ Characteristics

As shown in table 1, a total of 10 pediatric cancer patients participated in this study.

Table 1. Demographic characteristics of the participants

Pediatric patients (N=10)		**Overall N (%)**
Age (mean (SD))		8.7(1.8)
Gender		
	Male	3 (30)
	Female	7 (70)
Reason for visiting the hospital*		
	Pediatric ED	2 (15)
	Outpatient	9 (70)
	Inpatient	2 (15)

* The respondents could select multiple responses here

3.2. Conversation analysis

3.2.1. Used data log

There were a total of 1,676 interaction logs from 10 participants in this study, with 1,003 logs stored in ADMIN (Robot) and 674 logs stored in USER (Child). Among them, the

robot utilized a total of 955 logs and the child used a total of 332 logs, except for logs that were dropped due to network errors and research environment problems.

3.2.2. The delay time of the interaction log

The interaction log data delay time between the robot and the child is 5.01 seconds (sec) on average. In the case of the robot, there was a difference of at least 1.88 sec to at most 9.17 sec, with an average delay time of 4.29 sec. In the case of children, there was a difference of at least 2 sec to at most 11.96 sec, with an average delay time of 7.2 sec, confirming that the interaction log data of a child was stored more slowly than a robot for more than 2 sec.

3.2.3. Sentence similarity analysis

As shown in table 2, a total of 1289 log data were used. It was collected 97.2% accurately for log data of robot and 46.2% for log data of child.

Table 2. The similarity of real interaction and log data of robot and child by using TF-IDF (Term Frequency-Inverse Document Frequency)

				N (%)					
	N of log data			ROBOT			CHILD		
Subject	Total	One letter	Used log data	Same	Similar	Different	Same	Similar	Different
S1	103	4	99	83 (97.65)	2 (2.35)	0 (0)	10 (71.43)	2 (14.29)	2 (14.29)
S2	106	7	99	70 (98.59)	1 (1.41)	0 (0)	16 (57.14)	10 (35.71)	2 (7.14)
S3	96	2	94	71 (100)	0 (0)	0 (0)	15 (65.22)	3 (13.04)	5 (21.74)
S4	185	5	180	110 (98.21)	1 (0.89)	1 (0.89)	48 (70.59)	17 (25)	3 (4.41)
S5	204	7	197	132 (99.25)	1 (0.75)	0 (0)	28 (43.75)	29 (45.31)	7 (10.94)
S6	96	3	93	92 (98.92)	1 (1.08)	0 (0)	0 (0)	0 (0)	0 (0)
S7	107	2	105	73 (100)	0 (0)	0 (0)	13 (40.62)	14 (43.75)	5 (15.62)
S8	165	4	161	127 (99.22)	0 (0)	1 (0.78)	9 (27.27)	17 (51.52)	7 (21.21)
S9	136	2	134	106 (99.07)	1 (0.93)	0 (0)	15 (55.56)	10 (37.04)	2 (7.41)
S10	129	2	127	98 (97.03)	1 (0.99)	2 (1.98)	8 (30.77)	14 (53.85)	4 (15.38)
Total	**1327**	**38**	**1289**	**962**(97.2)	**8**(0.85)	**4**(0.35)	**162**(46.2)	**116**(32)	**37**(11.8)

3.3. Sentiment analysis

As a result of Sentiment analysis based on children's interactions, it was confirmed that 73% of the total interaction time was Neutral, 13.59% Positive, and 12.42% Negative. Additionally, evaluation according to the observation of pediatric psychologists

confirmed that they have curiosity (n=7, 70.0%) as the first response to the average robot. The overall responses and attitudes of the interaction were active (n=5, 50.0%) and passive (n=5, 50.0%), sympathetic (n=7, 70%), good at concentrating on the interaction (n=6,60.0%), and highly evaluated for interest (n=5, 50.0%). Finally, the overall emotions of the participants were positive (n=9, 90.0%), and the initiative of the interaction was evaluated as low (n=6, 60.0%).

3.4. Interview

After the interaction, children were interviewed in-depth about their experience with the robot. Under the Like category, the robot was perceived as a good communication partner and emotional supporter, with natural and mutual interactions. Under the Dislike category, the robot's interaction and response speed were perceived as slow, with interruptions in the middle of the interaction. Under the Wants category, children expressed a desire for the robot to improve its interaction speed, accuracy, and content diversity.

4. Discussion

The study has the following limitations: 1) The data were collected from a limited number of patients in a single institution, making it difficult to generalize the findings. 2) The study involved short-term interactions, and therefore, a long-term study of the interactions is necessary to derive more robust conclusions. 3) One of the child's log data (S06) was omitted from the DB due to a network error. However, as a preliminary study introducing a care robot for pediatric patients, this research is of great significance as it provides insight into the feasibility and necessary factors for future research, including trial and error analysis.

5. Conclusions

In this study, the degree of interaction between the SR and the child was explored using the interaction log collected in real-time, and the child's attitude toward the robot was confirmed. In addition, there was a difference in the attitude of interaction according to the characteristics, and positive emotions were confirmed in common. Accordingly, it is necessary to improve the network environment and optimize the technology protocol for data collection to improve log recording completeness in a real-time environment. In addition, if you interact using conversations that reflect the characteristics of children, the quality of emotional analysis is expected to improve.

References

[1] Torta E, et al. Evaluation of a small socially-assistive humanoid robot in intelligent homes for the care of the elderly. Journal of Intelligent & Robotic Systems. 2014 Sep;76:57-71.

[2] Szabóová M, Sarnovský M, Krešňáková VM, Machová K. Emotion Analysis in Human–Robot Interaction. Electronics. 2020;9(11):1761.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230230

Methodology for the Description of Socio-Technical Systems: A Case Study Approach

Myriam LIPPRANDT[a,1], Andrea D. KLAUSEN[a] and Rainer RÖHRIG[a]
on behalf of the DESIREE Study Group
[a] *Institut of Medical Informatics, Medical Faculty of RWTH University Aachen, Aachen, Germany*
ORCiD ID: Myriam Lipprandt: 0000-0001-9371-0551,
Rainer Röhrig: 0000-0002-0032-5118,

Abstract. The ethical implications and regulatory requirements of AI applications and decision support systems are generally the subjects of interdisciplinary research. Case studies are a suitable means to prepare AI applications and clinical decision support systems for research. This paper proposes an approach that describes a procedure model and a categorization of the contents of cases for socio-technical systems. The developed methodology was applied to three cases and serve the researchers in the DESIREE research project as a basis for qualitative research and for ethical, social, and regulatory analyses.

Keywords. bioethical issues, health technology assessment, socio-technical system, privacy, clinical decision support systems, telemedicine

1. Introduction

The digital transformation in healthcare is driven by applications such as telemedicine, clinical decision support systems (CDSS), and novel AI applications [1]. These applications have an impact on the doctor-patient relationship and the patient's social environment [2]. Normative challenges have risen in terms of responsibility, privacy, security, and autonomy. Furthermore, human-machine interaction and professional self-image are concerned. In the DESIREE project (https://www.desiree-forschung.de/desiree/), the ethical and social implications of CDSS were investigated. Three exemplary cases were used: nephrology, surgery, and nursing. The three cases were analysed with the aim of systematically describing the socio-technical and ethical challenges and the "side effects" of these applications. In our research, we tried to be aware, that there are always intended and unintended effects in social action. In our analyses, we focused also on the unintended effects. Then these more grounded descriptions were used by the interviewers for the interviews they conducted with the interviewees: patients, medical students, and nursing students on the impact of digitalisation.

[1] Corresponding Author: Myriam Lipprandt, Institute of Medical Informatics, University Hospital RWTH Aachen, Pauwelsstr. 30, 52074 Aachen. E-mail: mlipprandt@ukaachen.de

The aim of this reflected research was to develop a more grounded catalogue of criteria to follow and systematically describe AI, CDSS applications in terms of its socio-technical implications.

2. Method

Criteria underlying decision support systems were identified in an interdisciplinary team consisting of physicians, ethicists, and medical informaticians based on the methods of case study research according to Yin RK (2009) [3]. In the first step we asked the questions "how" and "why" according to the proceeding in case studies [3]. The interdisciplinarity consisted of identifying and analysing the entire socio-technical context. Here, the direct effects of human-machine interaction and indirect effects on professional self-image, legal requirements, and the effect on patients must be considered. Therefore, first, the direct interaction of the system was described and then, second, social effects were identified. In an iterative process, the procedure, the resulting categories, and their items were developed.

3. Results

3.1. Identifying the categories

The structured interdisciplinary procedure resulted in a categorization of cases for socio-technical systems. The following figure 1 shows the process of describing the case studies. The result was a set of categories to be adopted to all the cases:

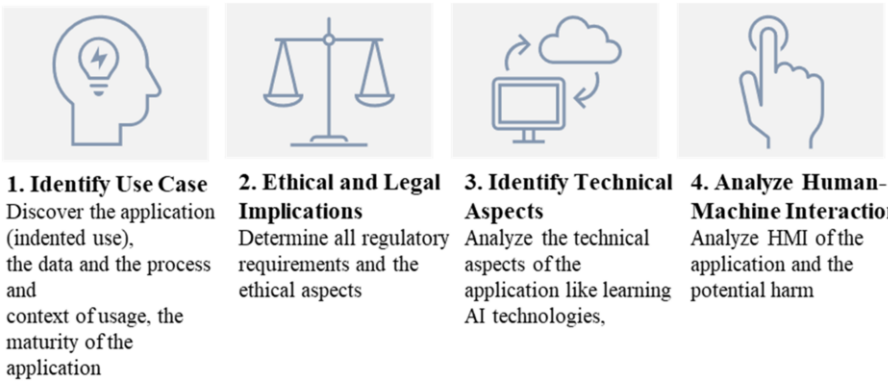


Figure 1. Process of categorization in establishing case studies for socio-technical systems

The figure 1 shows the process of identifying and categorizing the cases. The first step is identifying the application depending on the research question. In the DESIREE project, research projects from different medical domains were chosen. It is crucial to get detailed information about the product/research project. In the second step, all regulatory requirements and ethical implications have to be assessed. In the case of research projects, often there are no regulatory requirements easily to identify yet; here, experts have to classify the system and identify the potential regulatory framework. In

the third step, particularly in the case of AI, a distinction must be made between self-learning systems, data sources, and whether the system could also be implemented at another institution. For example, at another institution it may be dependent on the specification of the hardware. In the fourth and final step, interactions and hazards as well as possible usage errors and failures in the human-machine interaction must be assessed.

3.2. Categories

The category ***Use Case*** describes the overall system with its intended purpose. If it is not a medical device or purely a prototype from research, the intended use in clinical research is often not explicitly given. Here it is helpful to ask the researchers directly.

- **Intended purpose**: The intended purpose is a regulatory term from the Medical Device Regulation (MDR 745/2017).
- **Medical Domain:** This category describes in a nutshell the overall CDSS
- **Matureness** If the CDSS is not a medical device it could be an idea or a prototype from research
- **Goal of the System:** high level description of the purpose if there is no intended purpose described

The ***Ethical implications*** are ethics and social aspects to be discussed:

- **Ethical dimension**, e.g. responsibility, autonomy, transparency, trust, privacy, justice/fairness, caring
- **Decision and data transparency** raise the question of whether the decisions of the applications are clear and transparent to the user. Does the user know on which data basis the application is relying? Does the user know with which rules and above all reliability the system works, and the final results are achieved?
- **Impact on role and self-perception** bring up the question of whether CDSS change self-perception in terms of competence and self-image through applications that may be "better" than human decision making.
- **Impact on doctor-patient relationship** raises the question of whether CDSS change the doctor-patient relationship in terms of competence of the application and how it is perceived in the doctor-patient relationship.
- **Educational Impact** Are CDSS changing the way we need to educate doctors and nurses?
- **Impact on employment structures**: Will CDSS change the way we work in health facilities?

Legal implications refer to the legal and regulatory requirements.

- **Regulatory Aspects** (e.g. Medical Device Regulation): The development of medical devices is subject to regulatory requirements. The general safety and performance requirements must be fulfilled (addresses patient safety).
- **Privacy**: Are personal data (mostly those of the patient) sufficiently protected according to the law?
- **Performance** means the ability of a product to fulfil the intended aim or stated goal of the CDSS;
- **Risks**: What risks can the CDSS pose to humans, animals, and the environment? How high is the harm and the occurrence of an adverse event caused by the CDSS?

The ***technical aspects*** describe everything about the technology of the CDSS.

- **Method** describes the methodological basis on which the CDSS was developed. Whether it is a self-learning AI or contains a guideline-based rule set.
- **IT-Security, Data Protection** describes whether the data is safe e.g. from attacks.
- **Used Technology** describes whether the CDSS is an app or a desktop application (e.g. webservice) and whether there are other technological aspects that are relevant.
- **Components** such as hardware or additional systems that interact with or are included.
- **Users:** the ones who interact directly with the system. (e.g. medical staff, patients)

Human-machine interaction is a relevant category because ethical aspects and e.g. risks can be derived from the interaction.

- **Representational Layer**: the GUI but also the presentation of the data and the design are shown here.
- **Use Scenarios/User Story** describes the process of interaction in steps to clarify the interaction.
- **Use errors, Risks, Hazardous situations**: Based on the use scenarios, use errors and their effects can be determined.

3.3. Conclusions out of the three case studies

In the DESIREE project there were three cases, where we applied the developed categories: Case 1 describes the MeSiB system to give people in home ventilation more security through a safety box in case of disconnection of the ventilation tube or power failures. Case 2 describes an app to help physicians create personalized treatments for patients with kidney disease (https://ckdn.app/). Case 3 "Surgery of the future" is a support system designed to assist the surgeon in making the correct incision during an surgery.

All three cases were research projects. In order to work on all criteria, direct contact with the researcher or manufacturer is needed. Since all contacts to the researchers were available in this project, it was possible to apply all categories. Without this contact or a real user, the concrete description of the human-machine interface and underlying technology is not possible. Moreover, the impact of these systems is so diverse that an interdisciplinary team should work on the cases. Otherwise, there is a danger that a certain bias will occur. Not only interviewees, but, also interviewers (from different scientific fields) tend to stress risks, e.g., concerning data privacy and data protection.

4. Discussion

The systematic case study of three cases in the DESIREE project aims to describe the use of a CDSS more grounded in the context of regulation, ethics, and human-machine interaction. The methodology for the description should promote the understanding of the CDSS and serve as a basis for the development of a vignette for scientific investigations in the fields of social research, ethics, and innovation research.

The ethical aspects are a central part of this research. The regulatory aspects address many ethical dimensions such as responsibility or autonomy. Furthermore, it could be possible to include other ethical assessments such Meestar [4].

The legal and regulatory aspects such as the Medical Device Regulation (MDR), patient safety, and risks were considered crucial for a comprehensive picture of a socio-technical system. The legal frameworks are deeply linked to the ethical considerations.

This interaction is often safety-critical and is closely linked to acceptance criteria but also contain hazards for users. These hazards often relate to use errors, which should indicate what happens in the event of an error. But, often products only show the "happy path" and in research projects, there is often no risk management according to ISO 14971 and DIN EN 62366.

This research looked into three different cases resulting in a more in-depth view of the respective CDSS or AI application. Thus it should serve as a basis for further socio-technical research.

Contributions of the authors

ML: conception, design, writing, data analysis and interpretation, ADK: data collection, analysis and interpretation, revision, RR: data interpretation, revision. The DESIREE study group are from Fraunhofer-Institut für System- und Innovationsforschung ISI, Karlsruhe Tanja Bratan, Heike Aichinger, Nils Heyen, Diana Schneider. Institut für Ethik, Geschichte und Philosophie der Medizin, Medizinische Hochschule Hannover Sabine Salloch, Florian Funer. Evangelische Hochschule Rheinland-Westfalen-Lippe Martin Langanke, Wenke Liedtke

Acknowledgment

The DESIREE project was funded by the German Federal Ministry of Education and Research (BMBF) (funding code: 1GP1911D).

References

[1] Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ digital medicine. 2020 Feb 6;3(1):17.

[2] Čartolovni A, Tomičić A, Mosler EL. Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. International Journal of Medical Informatics. 2022 May 1;161:104738.

[3] Yin R K 2009 Case study research: Design and methods (Applied social research methods series vol 5) 4th edn (Los Angeles, Calif.: Sage)

[4] Manzeschke A. MEESTAR: ein Modell angewandter Ethik im Bereich assistiver Technologien. Technisierung des Alltags. Beitrag für ein gutes Leben. 2015:263-83.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230231

How Do User Participation and IT Self-Efficacy Influence User Attitudes Towards Smart Hospital Technology?

Saskia KRÖNER[a,1], Jörg HASSMANN[a], Moritz ESDAR[a,] Julia MAISCHAK[a] and Ursula HÜBNER[a]

[a] *Health Informatics Research Group, University AS Osnabrück, Germany*

Abstract. Smart hospitals aim to advance digitalization to provide better and safer care and increase user satisfaction by minimizing documentation burden. The aim of this study is to investigate the potential impact and its logic of user participation and self-efficacy on the pre-usage attitude and behavioural intention towards IT for smart barcode scanner-based workflows. A cross-sectional survey was conducted in a system of 10 hospitals in Germany that are in the process of implementing intelligent workflow technology. Based on the answers of 310 clinicians, a partial least squares (PLS) model was developed which explained 71.3% of the variance in pre-usage attitude and 49.4% of the variance in behavioural intention. User participation significantly determined pre-usage attitude through perceived usefulness and trust, while self-efficacy significantly did so through effort expectancy. This pre-usage model sheds light on how users' behavioural intention towards using smart workflow technology could be shaped. It will be complemented by a post-usage model according to the two-stage model of Information System Continuance.

Keywords. Smart hospital, user participation, self-efficacy, intelligent workflow, documentation burden

1. Introduction

Smart hospitals aim at designing workflows that are defined by seamlessness and intelligence and that are supporting comprehensive information logistics. The concept of a smart hospital should help healthcare professionals focus on better and safer care while increasing user satisfaction by, for example, minimising documentation and administration burden [1]. Putting this concept in place is a demanding long-term organisational undertaking that needs to bring together technology and people from early on. Building on the well-studied role of user participation [2], it can be expected that extensive involvement of the clinical workforce allows the development of smart technology to learn from human experience. Alike, it is the aim of user participation to develop a sense of ownership among the users and hereby increase their acceptance and motivation to use the technology. In addition, user participation is expected to generate a number of management-related, methodological, and cultural benefits [3].

[1] Corresponding Author: Saskia Kröner, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-mail: s.kroener@hs-osnabrueck.de.

However, there are more factors fostering the behavioural intention and use of smart technology as suggested by acceptance theories that emphasize the role of perceived usefulness and effort expectancy [4]. Modifications of these theories extended them by a variety of other factors characterising the user. They include IT self-efficacy [5] which is defined as one's own capacity to know that one can handle IT successfully. It has been shown that individuals with a high level of self-efficacy are more likely to facilitate goal achievement and that perceived overall self-efficacy contributes significantly to the motivation and performance of the individual [6].

The aim of the overall study is to better understand underlying mechanisms that influence the behavioural intention and use of technology and applications for smart workflows. In this part of the study, we sought to identify relevant factors determining the pre-usage attitude and intention to use the system in a pre-usage stage. The research questions, therefore, asked (i) whether user participation and self-efficacy can make a difference in positively influencing the clinical workforce's pre-usage attitude and if yes (ii) what the underlying mechanisms are that can explain this influence.

2. Methods

2.1. Project description and user participation

To pursue this aim, a long-term formative evaluation study was initiated to accompany a group of hospitals on their way towards becoming a smart hospital. The Klinikum Region Hannover (KRH) is a system of seven general hospitals and three special hospitals in the Hannover region (Lower Saxony, Germany). KRH has approx. 8,500 employees, 3,400 beds and 100,000 cases per year. KRH's activities to implement smart workflows were bundled in the "ScanProCare!" project to automate routine activities and intelligently guide user management in selected processes. Twenty-six processes and tasks related to clinical materials management and documentation were chosen as examples of highly relevant processes by the KRH management that can be standardized and automated across the hospital system. The clinical materials management workflows comprised amongst others "material requests" and "inventory management", clinical documentation processes included "material consumption during an operation" and "implants used". Information embedded in the various processes was linked through the patient/visit ID.

"ScanProCare!" was launched by six one-day workshops. One of them on 30th Sept 2019 with 15 employees from the administration, five partly parallel sessions on 16th/30th Jan and 27th Feb 2020 with 55 clinical key users of the hospital units concerned (operating room, endoscopy, radiology, and cardiac catheterization laboratory (20), normal ward and intensive care units (21), delivery room and anesthesia (14)). The workshops were organized by KRH and moderated by two KRH managers, two researchers of Hochschule Osnabrück and two software engineers from the IT vendor. The goal was to identify (i) the current practice of the selected processes in the various units, (ii) further processes that should be considered in the project and to herewith lay the foundation (iii) of the requirement concept of the different project modules. In addition, all selected processes were studied by interviewing 25 clinical experts in all units concerned in four hospitals for two days each while observing them performing their tasks. The observations and the resulting process modelling were performed by a researcher from Hochschule Osnabrück

Software development, installation and rollout are conducted by GSG mbH Hannover Germany. The digital workflows enabled by the software are controlled using a set of rules in a central intelligent hub. It triggers the application systems to deploy messages intelligently to the subsystems and to capture the data at the point of care in the sense of smart information logistics. The end-users are working with an android-based mobile device – similar to a smartphone with an additional in-built barcode scanner. The software was implemented in two pilot institutions and accompanied by on-site training of the users. User feedback concerning the functionality and usability of the software app was immediately incorporated in an agile manner.

In summary, user participation was enabled through the six key user workshops with 70 key users, 25 expert interviews combined with the process observations and an agile programming style at the two pilot sites.

2.2. Data collection and conceptual model

A cross-sectional web survey methodology was employed to capture the data for answering the research questions. To this end, a questionnaire with 31 closed questions on the constructs *user participation*, *self-efficacy*, *perceived usefulness*, *effort expectancy*, *social influence* (covariate), *facilitating conditions* (covariate), *trust*, *pre-usage attitude* and *behavioural intention* was designed in accordance with the expanded two-stage model of Information System Continuance [7], which also includes the core constructs of the unified theory of acceptance and use of technology (UTAUT). The target group was the clinical staff in the relevant units of KRH (n = 4742) irrespectivly of wether they would use the system in the immediate or later in the far future. 310 participants (response rate = 6.5 %) returned a complete questionnaire. The survey took place anonymously and voluntarily in the period from May 2021 to September 2021. Figure 1 shows the conceptual model underlying the questionnaire with a total of 9 constructs. We applied structural equation modelling (SEM) to test the interrelationships between the constructs. Partial least squares structural equation modeling (PLS-SEM) was used to test the effects of *user participation* and s*elf efficacy* on the pre-usage belief and the attitude towards using barcode scanners.

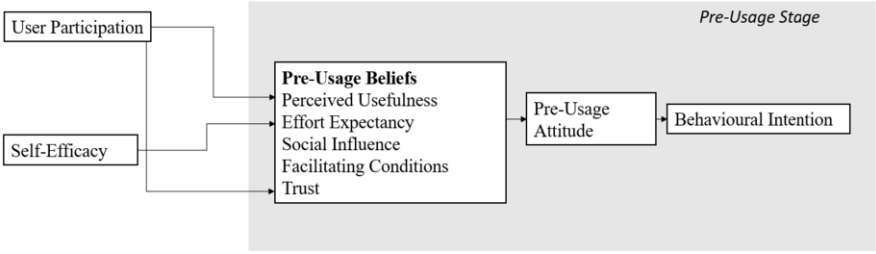


Figure 1. Conceptual model

3. Results

The sample consisted of data from 310 participants (female: 74.8%; male: 24.5%; divers: 0.7%). The median age was 47.3 years and median professional experience was 23.3 years. The occupational group was composed of nurses (78.7%), (medical) technical assistants (7.3%), and others (14.0%).

The structural equation model (Figure 2) explained 71.3% of the variance in *pre-usage attitude*, 49.4% of the variance in *behavioral intention*, and 35.5% of the variance in *effort expectancy* (Fig. 2). *Self-efficacy* and *user participation* were found to exert a positive influence. *Self-efficacy* influenced *effort expectancy* (β=0.596) and *user participation* determined both *trust* (β=0.264) and *perceived usefulness* (β=0.203). *Pre-usage attitude* was influenced by *trust* (β=0.388), *perceived usefulness* (β=0.273), *social influence* (β=0.209) and *effort expectancy* (β=0.178). It in turn determined the *behavioural intention* to use the scanner (β=0.570). Contrary to our initial assumptions, there was no significant effect of the *facilitating conditions* on the *pre-usage attitude* towards using the technology.

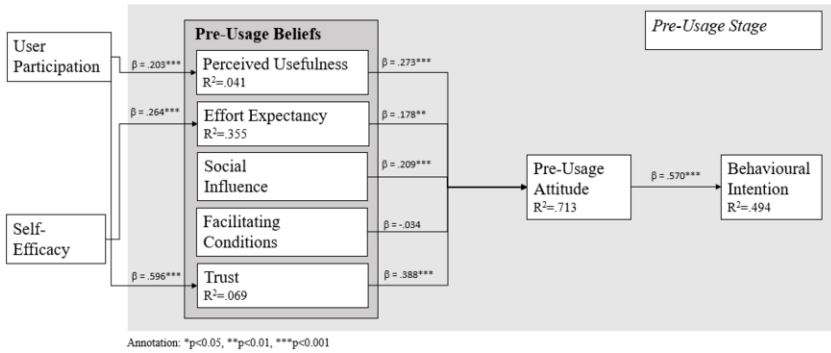


Figure 2. Structural equation model

All parameters assessing the measurement models pointed to valid specifications of the reflective models as well as the formative model in terms of convergent validity and internal consistency. In addition, sufficient discriminant validity was established according to the heterotrait-monotrait ratio of correlations. Also, no collinearity was found in the structural model, as all the inner variance inflation factor values ranged within the limits of 0.20 and 4. This model thus meets all requirements and can be regarded as formally valid.

4. Discussion and conclusion

The model proposed is able to explain a large percentage of the variance in *pre-usage attitude* and *behavioural intention*, the two main dependent variables. From an input perspective *user participation* shaped the *perceived usefulness* of the application and the *trust* that users laid in the system, whereas *self-efficacy* determined *effort expectancy*. With regard to research question (i), it can be summarized that both *user participation* and *self-efficacy* positively influence the *pre-usage attitude* towards the new technology. While *user participation* exerts this influence by building *trust* and allowing the users to develop an idea about the usefulness of the system (*perceived usefulness*), *self-efficacy* facilitates *effort expectancy*. It is through these mechanisms (research question (ii)) that *user participation* and *self-efficacy* promote the *pre-usage attitude* and eventually the *behavioural intention*.

The strong impact of user participation revealed in this study reflects the variety of measures taken in this project to involve the users. The involvement embraced explaining the need for the technology as well as giving users the opportunity to define the target

processes and how the technology should ease their work. The active role they played combined with deep insights given to the users equipped them with an understanding and trust in the "why" and "how". The fact that user participation is paramount is not new [8,9], however, the explanations found in this study add further substance to its logic.

While user participation is an active measure at the organisational level, self-efficacy that determines effort expectancy is a characteristic of the individual. Its role regarding technology acceptance is not new [10] - alike user participation, however, this study contributes to revealing the underlying mechanism. Contrary to our expectations and prior research [4], facilitating conditions did not affect acceptance in this study.

The low response rate may be due to the fact that only persons responded who felt that this project had an immediate impact on their work. As a pre-usage concept, this model cannot explain the actual use of the technology. Its value for the entire technology adoption process is therefore limited and has to be extended regarding the post-implementation and usage phase as proposed by the Information System Continuance model [7]. This part of the overall study is a work in progress: the post-implementation questionnaire has been recently deployed to the clinicians who now actually use the system. Modelling the post-implementation situation at various points of time and hereby stepwise capturing behavioural usage of the technology will render a complete picture of smart workflows in a future smart hospital.

Acknowledgement

This study was funded by the EU and the State of Lower Saxony (EFRE grant no. ZW 3- 85033011). We wish to thank our project partners GSG mbH (Marc Elmhorst, Martin Stache, Elias Sakkal) and Klinikum Region Hannover GmbH (Barbara Schulte, Tim Bauer, Vanessa Herbst) for providing the support necessary to conduct this study.

References

[1] Kwon H, An S, Lee HY, Cha WC, Kim S, Cho M, Kong HJ. Review of Smart Hospital Services in Real Healthcare Environments. Healthc Inform Res. 2022 Jan;28(1):3-15. doi: 10.4258/hir.2022.28.1.3

[2] Hartwick J, Barki H. Explaining the Role of User Participation in Information System Use. Management Science 1994 Apr;40(4):440-465. doi: 10.1287/mnsc.40.4.440

[3] He J, King WR. The role of user participation in Information systems development, implications from a meta-analysis, J of Manag Inf Sy 2008,25 (1):301–331.

[4] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003, 27:425–478.

[5] Compeau DR, Higgins CA. Computer Self-Efficacy: Development of a Measure and Initial Test. MIS Quarterly 1995 Jun,19(2):189-211. doi.: 10.2307/249688

[6] Bandura A, Locke E. Negative self-efficacy and goals effects revisited. Journal of Applied Psychology, 2003;88(1):87–99.

[7] Venkatesh V, Thong JYL, Chan FKY, Hu PJ-H, Brown SA. Extending the two-stage information systems continuance model: incorporating UTAUT predictors and the role of context. Information Systems Journal 2011;21(6):527-555. doi:10.1111/j.1365-2575.2011.00373.x

[8] Kujala S. User involvement: a review of the benefits and challenges. Behaviour & Information technology 2003;22(1):1-16.

[9] Bano M, Zowghi D. A systematic review on the relationship between user involvement and system success. Information and Software Technology 2015 Feb;58:148-169.

[10] Mensah IK, Zeng G, Mwakapesa DS. The behavioral intention to adopt mobile health services: The moderating impact of mobile self-efficacy. Front Public Health. 2022 Sep 27;10:1020474. doi: 10.3389/fpubh.2022.1020474.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230232

Improving Well-Being in Schools - Lessons Learned from IoT Experts

Agnieszka KITKOWSKA[a,1] and Karin AHLIN[a]
[a]*Karlstad University, Sweden*
ORCiD ID: Agnieszka Kitkowska https://orcid.org/0000-0001-7384-4552
ORCiD ID: Karin Ahlin https://orcid.org/0000-0003-4051-6960

Abstract. The Internet of Things (IoT) is a source of knowledge about the surrounding environment and people in such an environment. The insights collected by IoT can provide the knowledge needed to improve people's health and overall well-being. Schools are one environment where IoT is scarcely applied, yet, it is expected that this is where children and teenagers spend most of their time. Drawing on previous findings, this paper presents preliminary results from qualitative inquiry investigating how and what IoT-based solutions could support health and well-being in elementary educational settings.

Keywords. IoT, Well-being, Expert interviews, User requirements, Health

1. Introduction

Children, including teenagers, typically spend most of their time at the school premises, and their educational performance, health, and well-being may be affected by the school environment [1]. Moreover, the school environment also affects the well-being of teachers and other school staff. Therefore, ensuring that the school environment supports flourishing well-being is essential [2].

Technologies such as the Internet of Things can assess the quality of the school environment. The ubiquitous IoT sensors can measure indoor quality indicators, e.g., oxygen level, dust particles, CO2, and temperature, to name a few. IoT might also be used to assess the well-being and health of individuals using more (e.g., wearables) or less (e.g., infrared cameras) invasive technologies. Despite the availability of these technologies in the market, it is uncommon for schools to apply them. One reason is that we need more insights on how and what IoT-based solutions could help to fulfill the needs or requirements of school staff and students.

In this research, we aim to gain such insights by asking, *How can IoT improve the well-being of students and staff?* To answer this question, we use a qualitative inquiry based on the previous research's findings, where we identified user requirements[2] defined in the case of a Swedish school[3]. We engaged with IoT experts to identify

[1] Corresponding Author: Agnieszka Kitkowska, E-mail: agnieszka.kitkowska@kau.se.

[2] The research is part of a larger project IoT in Schools—collaborative project building IoT testbed in school, engaging industry and public sector.

[3] The school was for students of 7th-9th grade. Interested reader can find more information about the Swedish school system on https://www.norden.org/en/info-norden/compulsory-schools-sweden.

possible scenarios of using IoT to build solutions that aim at improving health and well-being in the context of elementary school. The reported results give preliminary insights into the collected data.

2. Method

The expert interviews were used as a method for qualitative inquiry. Eleven experts working in the industry and academia participated. All experts had IT backgrounds and worked as privacy and security specialists, IoT engineers, or in managerial and sales positions in IoT companies. The interviews were conducted online, recorded, and transcribed. Each interview was approximately an hour long. Six interviewees were Swedish, and others were from Norway, Spain, Germany, the UK, and the USA.

The expert interview is one of the qualitative methods which focuses on the interviewees' knowledge—their expertise in a given field [3]. We chose expert interviews for two reasons. First, because the study is part of a larger project and in the previous studies, we discovered that generally, people have very little or no knowledge about IoT. Second, the interviews were designed based on the findings from our previous studies, where we identified user requirements based on the needs of the school staff (teachers and healthcare personnel) and students. We assumed that IoT experts, understanding IoT's technological foundations, can relate better to user requirements and shed light on potential IoT-based solutions that could be applied in school settings.

We choose the semi-structured interview approach to ensure a space for conversations between the interviewer and experts instead the strict format of structured interviews or overly loose conversations when using unstructured interviews [4]. The previously identified user requirements were core for the interviews. The interview contained three questions. First, we asked participants to briefly explain their relationship with IoT, focusing on their professional role. Second, we presented participants with user requirements (**Figure 1**) and asked them whether they thought these requirements (some or all) could be fulfilled with any IoT-based solutions. We instructed participants that they could think of their "dream solutions" and discuss not only the existing technologies but also future solutions yet to come. This part of the interview was the least structured and often led to a dialogue between the interviewer and experts. Lastly, we asked participants whether they could think of any barriers to applying IoT-based solutions in school settings and how to overcome them.

2.1. Data analysis

The interview transcripts were analyzed independently by two reviewers. Thematic analysis was used to identify potential patterns forming in the dataset. We used reflexive thematic analysis, following the steps recommended by Braun et al. [5]. First, researchers familiarized themselves with the data by reading through the transcripts. Next, they generated codes applying inductive orientation and creating semantic and latent codes. Then researchers discussed constructed themes. Lastly, they revised and defined themes in an iterative process. The reported results are preliminary and focus only on the topic of well-being, omitting barriers to IoT-based systems in schools. A few central themes were identified in the data set: Improvement of learning conditions, Improvement of teaching staff conditions, Safety of the school environment, Waste management and cleaning, and Data aggregation.

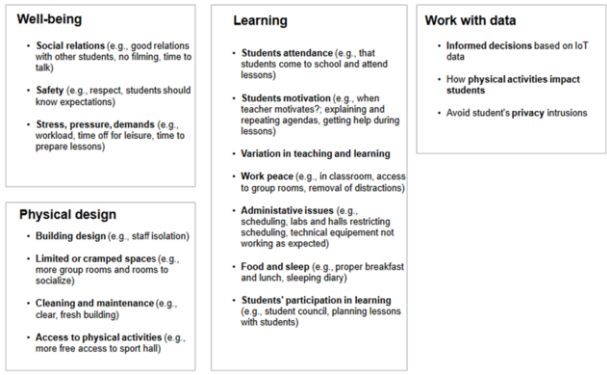


Figure 1. User requirements chart used during the interviews.

3. Results

Improvement of learning conditions. Experts discussed how IoT-based solutions could help students achieve better learning outcomes. They discussed how sensors measuring environmental changes, e.g., temperature, sound, and air quality, could be adjusted to ensure classroom well-being. They also discussed how IoT could help students learn programming or hacking devices implemented in school, preparing them better for their future digital life. Some mentioned how IoT could help detect learning issues just in time, e.g., by installing feedback devices in desks to enable students to ask questions anonymously. Teachers could receive questions immediately and address them. Such a solution could benefit students lacking self-confidence or who have learning difficulties. "You have a little button that you can press when you are confused. Well, obviously, you [...] might not want to be the one that says: I don't understand, and you know how it works. Like you always want someone else to be the one."

Improvement of teaching staff conditions. Experts mentioned how IoT could help teachers with assessing their own performance. E.g., microphones tracking changes in the voice tone could help teachers detect whether they are monotonous during lessons. Another idea was to use microphones to recognize how the teacher asks questions: is it a way that leads to pupils' engagement, or the teacher asks yes or no type questions? Others talked about creating better conditions for teachers through adjustments based on environmental sensors, as mentioned in the context of students. Similarly, sensors such as RFID could be used to measure attendance or automated adjustments of lightning in the classroom, potentially relieving teachers of some of their classroom duties.

Safety of the school environment. Interviewees talked about safety from two different points of view. First, they referred to safety in the context of incident response. E.g., wearable sensors or infrared cameras could automatically detect whether something happened to a student, e.g., a student fell. They also mentioned alarm systems, e.g., IoT buttons, that could be pressed in distress. The second point of view was to use environmental sensors to help detect harmful substances in the building, e.g., if there are any changes in the air (e.g., gas) or if students would smoke cigarettes at school premises.

Waste management and cleaning. Although intuitively, there might not be a direct link between well-being and waste management, some experts discussed it in the context of the school cafeteria and problems that often occur considering the dietary habits of youngsters. IoT could measure students' food consumption and inform the

kitchen about food popularity. Sensors could also help the cafeteria staff with recycling, e.g., cameras detecting where given items should be recycled, relieving the staff from the cognitive workload and potential stress. IoT-based solutions for cleaning and maintenance were obvious, e.g., cameras or sensors detecting dust particles or toxic substances in the air. Some suggested AI-based systems that cleaning companies could use. For instance, "I think IoT could play a vital role in like measuring air quality and it could also most likely just measure [...] how often the facilities are cleaned and maybe if you have some AI interpretation of what the cleaner looks like or [...] you could tag [the cleaner with something] that would, you know, be recognized by an AI solution."

Data aggregation. This theme was reoccurring among the experts' responses; IoT might not be sufficient to help ensure well-being in school settings. Its role is significant, but there is a need for algorithmic transformation of data and correlation of information coming from various IoT sensors. Other information, e.g., self-reported data, performance, food consumption, or external events that might affect students' behavior (e.g., sporting or gaming, which may affect students' sleep and, consequentially, school performance), should also be considered.

Levels of CO2 in a room could be correlated with students' cognitive performance to enable an understanding of what exact CO2 levels the classroom should have to ensure students' progress. Similarly, combining information about the number of students in the cafeteria with details of food consumption could be used to prevent students from leaving school premises to purchase takeaway; "[IoT can] measure the number of students that actually sit down and eat. [...] but you need to correlate the IoT data often with other data points like today it was lasagna, and we had 95% attendance from the students." Live information about the food through a system connecting with students' mobile phones or public displays could encourage students to stay in the school cafeteria, "in winter, we want […] the soup hot. […] I receive a signal that the soup is hot, [...] they receive the same message at the same moment, they would all run to get the hot food."

Experts mentioned how information from more sophisticated sensors measuring physiological responses could be used with machine learning to gain more insights into students' well-being. E.g., "So if you talk about future solutions, anything related to focus and even happiness, emotions, I think it would be very interesting if you could capture them and then take that data, crunch it to try to find some sort of meaning."

Correlated data can help with scheduling; e.g., indoor quality indicators combined with data from wearable sensors may predict when people should take a break and leave a classroom. "[I]f there are many people there and we can measure […] bad air and we can see that people's pulses, for example, increasing [...] we could just possibly [say] you need to get out of the room for 10 minutes and take a break."

4. Discussion

The interviewees gave good examples of approaching and designing IoT solutions to improve well-being. Only some of the requirements were left without any design proposal, implying that IoT is of interest to the school environment. The design varies concerning how data should be collected to meet the requirements. One example is the suggestion to correlate the information to help schedule attendance, e.g., RFID solutions. That the design varies implies a variation in implementations, requiring a high grade of knowledge about collecting data. Nowadays, the solutions are mainly stand-alone, meeting one or two similar requirements and failing to support student and staff well-

being fully. AI and ML could be utilized in systems integrating data from various sensors to present it in an understandable and useful manner. Although technically, such systems can be developed, their use might be constrained due to privacy and security risks, which are particularly important when processing data about vulnerable populations, such as minors. Currently, few IoT solutions focus on improving student and staff well-being. One reason could be the need for IoT knowledge within elementary schools or organizations supporting schools' IT solutions. The required knowledge is unique and limited to many IT staff working mainly in IT consultant companies. Another reason could be that well-being is subjective and, therefore, experienced as hard to measure objectively. Still, the findings show realistic ways to measure what is subjective, such as microphones assessing teachers' performance. Lastly, it might be that IoT suppliers do not focus on designing solutions within the school environment.

4.1. Limitations and future work

The present study was preliminary, being part of a larger project, and could be improved. E.g., we did not interview health professionals. However, the user requirements = used in the expert interviews were based on the school staff, including health personnel. Moreover, as the present research aimed to identify potential system design recommendations, non-tech professionals might lack an understanding of technology and are unlikely to provide significant contributions. Still, future work could target health professionals to assess their views on IoT-based systems for school well-being. Lastly, we applied an inductive approach that was suited to exploratory study. Future studies could apply a specific theoretical framework and deductive analysis.

5. Conclusions

Based on previously collected requirements, the IoT experts presented various ideas for designing IoT-based solutions and how data should be collected and aggregated to meet these user requirements. Nevertheless, the discussed solutions were stand-alone, adding to the complexity of using them to meet the requirements. We conclude that IoT suppliers see potential in designing IoT applications to increase the well-being of students and staff. Nevertheless, they still need to discuss how to present the data to decision-makers—staff, students, or caregivers.

References

[1] Blandford AE. Semi-structured qualitative studies. Interaction Design Foundation.
[2] Clement N. Student wellbeing at school: The actualization of values in education. International research handbook on values education and student wellbeing. 2010:37-62.
[3] Döringer S. 'The problem-centred expert interview'. Combining qualitative interviewing approaches for investigating implicit expert knowledge. International Journal of Social Research Methodology. 2021 May 4;24(3):265-78.
[4] Løhre A, Lydersen S, Vatten LJ. School wellbeing among children in grades 1-10. BMC public health. 2010 Dec;10(1):1-7.
[5] Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. The SAGE handbook of qualitative research in psychology. 2017 Jun 30;2:17-37.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230233

How Do Informaticians and IT-Architects Collaborate, or Not? A Case Study from a Public Health Care Provider

Anna ROSSANDER[a,1]
[a] *Department of Applied IT, University of Gothenburg, Sweden*
ORCiD ID: Anna Rossander https://orcid.org/0000-0002-7456-3346

Abstract. Despite years of work from both informaticians and IT-architects interoperability within healthcare is still low. This explorative case study performed on a well-staffed public health care provider shows that the involved roles were unclear, processes did not include each other, and that tooling was incompatible. However, interest in collaboration was high and technical advances and inhouse development were seen as incentives for increased collaboration.

Keywords. Health informatics; Informatician; IT-architect; Interoperability, Collaboration

1. Introduction

Today's installed base of health information systems leaves much to wish for regarding interoperability. Managing and developing such an ecosystem can be viewed as a type of infrastructuring work, and the work processes involved can be analysed with sociotechnical frameworks [1].

Two roles involved are informaticians and IT-architects, where the former develop standards and information structures and the latter curate the installed based including the accumulated technological debt [2,3]. Research has been done on EHR implementations highlighting that informaticians should work together with health care staff [4], but there is scarce literature on how informaticians and IT-architects should collaborate to leverage informatic work in health care information systems and thus increase interoperability.

This paper presents results from an explorative case study aiming to explore how these two key roles collaborate. The work is a response to frustration from practioners within informatics as well as IT, aiming to elucidate possible explanations to a real-world problem [5,6]

2. Method

A publicly funded and run regional health care provider covering a population of approximately 1,2 million inhabitants was used as case for this work. The organisation

[1] Corresponding Author: Anna Rossander, E-mail: anna.rossander@gu.se

had an IT-department catering for the entire organisation (including regional development, cultural activities and healthcare) with approximately 800 employees (whereof some are consultants). A health informatics unit was formed during 2018 and has since been staffed with between 20 and 30 informaticians.

Electronic health records (EHR) have been used in the organisation for approximately 30 years. The organisation was during data-collection going through configuration of a new central EHR system to replace several of the existing systems.

Primary data consists of nine semi-structured interviews which were made during November and December 2022. The informants were a purposive sample [7] of staff in different positions within the IT-department (n=6) and the health informatics unit (n=3). In addition to this, secondary data consisting of documents describing strategy, work processes and visions as well as informal talks have been used.

The interviews were semi structured and 45 to 90 minutes long. They were held and recorded via Teams and then transcribed by a professional transcriber. Bottom up open and selective coding as described by Urquhart [8] was done in ATLAS.ti [9]. The resulting codes were used for abductive reasoning. The interviews were made by the author who is a PhD-student and works at the regional health care provider.

3. Results

3.1. Stated Processes Does not Include 'the other'

The IT-department had a formal framework based on TOGAF enterprise architecture methodology and framework [10]. The regional adaption was well perceived amongst IT-architects, "*there are really good processes and ways of working with architecture in the region*." However, IT-architects and informaticians both stated that TOGAF and the regional adaption omitted informatics work

The informatics unit's processes were under development during data collection but at the time had no formal link to the IT-department.

3.2. Tools Do not Support Collaboration

The IT-department used the software iServer by Orbus Software and Visio by Microsoft to model, store and share documentation about architectural artifacts, whereas the informatics unit used Visual Paradigm for their information models. Neither had access to the other's system. This was perceived as a big hindrance "*the toolbox is a very large deficiency in the region… that one doesn't have unified tools do describe the same types of artifacts.*" leading to double work "*everyone in these [information] islands who work with the information… they have probably produced much material that never comes anyone else to gain."*

Informants from the IT-department were either unaware of what the informatics unit worked with or knew about their work but did not have access to it *"And I haven't seen what you have developed. How have you done it by the way? Have you done it on your own or together with…?* " This problem was perceived also from the informatics unit who wished for tools that could visualise their work; "*I believe much misunderstanding in such discussions on collaboration would have been avoided if the results [information model] would have been more accessible.*"

3.3. Unclear Roles

The IT-department had six described IT-architect roles (see table 1); five based on the Swedish IASAs roles [11] and an additional role called "domain architect". The "domain architect" had a similar strategic view as the enterprise architect but with a smaller scope covering only part of the enterprise.

Table 1. Architect roles

Role in IT-department	Present in Swedish IASA	Resembling global IASA
Enterprisearkitekt	Yes	Enterprise Architect
Verksamhetsarkitekt	Yes	Information Architect Business Architect
Lösningsarkitekt	Yes	Solution Architect
Mjukvaruarkitekt	Yes	Software Architect
Infrastrukturarkitekt	Yes	Infrastructure Architect
Domänarkitekt	No	-

The Swedish IASA roles were in turn based on the international IASA roles [12,13] (note: the international version has since been revised). During translation the international roles "information architect" and "business architect" have been merged to "verksamhetsarkitekt", which roughly translates to "business architect". Thus, neither the Swedish IASA nor the region had a role "information architect". "Information architect" is described by Swedish IASA as a specialisation of "verksamhetsarkitekt" "*focusing on those parts of architectural work concerning information management*" [11].

The knowledge of the stated roles within the organisation was varied among the informants from the IT-department and low among the informants from the health informatics unit. Some informants stated they had, or mentioned relation to, the role "integration architect" which is not included in the regional description of roles. This role is described by Swedish IASA as a "*software architect focusing on integrations between different systems*" [11].

IT-management perceived the roles as guidance and said that architects are expected to do work also within other roles, especially regarding informatics; "*Regardless of what type of architect role one has … it is part of one's basic competence to have some sort of idea about information architecture and informatic work at some level."*

The informants from the health informatics unit had little insight into the different types of architects at the IT-department. They introduced themselves as informaticians, but those who had medical background also used the title "medical advisor" to display their background towards healthcare staff. Some informaticians said they might as well be called "information architects", and that colleagues with similar jobs in other regions in Sweden used that title.

Informants from the IT-department were unsure what the role "verksamhetsarkitekt" meant and no informant could name anyone with that title. The informaticians had a common understanding of the role "verksamhetsarkitekt" as a person modelling a business workflow both for process improvement and as input to informatics work exploring what information was used in the business. This is a limited scope compared to the Swedish IASA role where also "*how information is managed including analyses of information quality*"[11] is included. The informaticians did not see themselves as "verksamhetsarkitekter", nor did they know of any such employed in the organisation.

It was also unclear what role, if any, could make decisions regarding choices of informatic standards, on direct question, "*Who sets common standards?*" one informant replied, "*Everyone and no one.*"

3.4. Interest in Collaboration High But not Perceived As Reciprocal

Informants from the IT-department showed interest in working with informaticians; "*..for several years [we] have been thinking about how we will interact with [informatics]*" but had the impression that this interest was not reciprocal "*No, the informaticians they… no, they don't want to hear from us. They don't want to talk to IT.*

The informaticians stated that the practical work done by colleagues at the IT-department mainly focused on integrations and data extraction for secondary use, and that there were few possibilities to work with structured information all the way from the user; "*..it [our work] hasn't been able to make a difference. Because it is more about the systems not working together.*" The informaticians showed frustration over this but had also chosen "*an approach closer to describing the business information in a system or application independent fashion"* and were organisationally closer to the health care department than the IT-department.

3.5. Window of Opportunity

Several informants indicated a window of opportunity for increasing collaboration. Informants from the IT-department referred to a recent regional strategic decision to develop more inhouse systems as one such facilitator; "*and now we have, for the first time in a long time, a chance to work proactively ... and also deepen collaboration with the informatics unit, to being also operative.*" The latter part of this quote refers to informatics work oftentimes not affecting documentation in use.

Informants from the informatics unit described the technical matureness of the installed base as a possibility to increase work with information structure; "*the more mature the technical architecture and technical settings ... the more incentive there is to look at the structure [of the data] to make sure that the data you send between systems can be compared informatically."* Informants repeatedly stressed the involvement of healthcare staff and stated that the technical matureness also was essential to involve such personnel; "*It is very hard to motivate people to make a big effort when we don't have capacity to make use of it."*

4. Discussion

Previous research has shown collaboration as a central quality of interoperability work within infrastructuring and has shown the need for both balancing of relations through sociotechnical negotiations and adequate tooling [14].

This work showed poor relations in the form of processes not including both groups and incompatible tooling. These hindrances for collaboration were known by both informaticians and IT-architects. One reason to why these known problems were not already solved might be the unclear roles that were a previously unknown problem. Social negotiations are hard to expediate when the roles involved are now known. Organisations should put effort into clarifying different roles involved in complex collaborative settings.

Informants from both groups were interested in improving collaboration, and saw increased, albeit different, openings for this. Incentives and hindrances regarding this will be explored further in forthcoming work.

The author's knowledge of the studied organisation is both a strength and a limitation of this work. Another limitation is the relatively small sample size of informants. Some results could be due to specificities in the studied case, but they might also be transferrable to other settings.

5. Conclusion

Collaboration between informaticians and IT-architects was poor. Roles were unclear, processes did not include each other, and tooling was incompatible.

However, interest in collaboration was high and technical advances and inhouse development were seen as incentives for increased collaboration.

References

[1] Hanseth O, Lyytinen K. Design theory for dynamic complexity in information infrastructures: the case of building internet. Journal of information technology. 2010 Mar;25(1):1-9.

[2] Lim E, Taksande N, Seaman C. A balancing act: What software practitioners have to say about technical debt. IEEE software. 2012 Aug 23;29(6):22-7. doi:10.1109/MS.2012.130.

[3] JMagnusson J, Juiz C, Gómez B, Bermejo B. Governing technology debt: beyond technical debt. InProceedings of the 2018 International Conference on Technical Debt 2018 May 27 (pp. 76-84). doi:10.1145/3194164.3194169.

[4] Eden R, Burton-Jones A, Ballantine C, Donovan R, McKavanagh D, Staib A, Sullivan C. The Digital Transformation Journey of a Large Australian Hospital: A Teaching Case. Communications of the Association for Information Systems. 2022;51(1):25. doi:10.17705/1CAIS.05134.

[5] Haux R, Kulikowski CA, Bakken S, de Lusignan S, Kimura M, Koch S, Mantas J, Maojo V, Marschollek M, Martin-Sanchez F, Moen A. Research Strategies for Biomedical and Health Informatics. Methods of Information in Medicine. 2017 Jan;56(S 01):e1-0. doi:10.3414/ME16-01-0125.

[6] Desouza KC, El Sawy OA, Galliers RD, Loebbecke C, Watson RT. Beyond rigor and relevance towards responsibility and reverberation: Information systems research that really matters. Communications of the Association for Information Systems. 2006 Mar 13;17(1):16. doi:10.17705/1CAIS.01716.

[7] Moser A, Korstjens I. Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. European journal of general practice. 2018 Jan 1;24(1):9-18. doi:10/gf362z.

[8] Urquhart C. Grounded theory for qualitative research: A practical guide. Sage; 2022 Sep 23.

[9] ATLAS.ti Scientific Software Development GmbH, (2022). https://atlasti.com.

[10] An Introduction to the TOGAF® Standard, 10th Edition, n.d. https://pubs.opengroup.org/architecture/w212/?_ga=2.67959121.1374010118.1673094148-912721077.1673094148 (accessed January 7, 2023).

[11] Värild J, Larsson A, Ogenstad C, Thor E, Padron PE. Arkitektroller för den digitaliserade organisationen, (2020). https://www.iasa.se/wp-content/uploads/2020/02/IASA-Arkitektroller-2020.pdf.

[12] Global I, BTABoK Competency Model 3.0, *BTABoK*. (2022). https://btabok.iasaglobal.org/btabok_3/capability-taxonomy-and-descriptions-3-0/ (accessed January 7, 2023).

[13] Hackney H, Adopting the BTABoK Competency Model, *Architecture & Governance Magazine*. (2022). https://www.architectureandgovernance.com/applications-technology/adopting-the-btabok-competency-model/.

[14] Ulriksen GH, Pedersen R, Ellingsen G. Infrastructuring in healthcare through the openEHR architecture. Computer Supported Cooperative Work (CSCW). 2017 Apr;26:33-69. doi:10.1007/s10606-017-9269-x.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230234

Ethical Perspectives on Implementing AI to Predict Mortality Risk in Emergency Department Patients: A Qualitative Study

Lena PETERSSON[a,1], Kalista VINCENT[a], Petra SVEDBERG[a], Jens M NYGREN[a] and Ingrid LARSSON[a]
[a] *School of Health and Welfare, Halmstad University, Halmstad, Sweden*

Abstract. Artificial intelligence (AI) is predicted to improve health care, increase efficiency and save time and recourses, especially in the context of emergency care where many critical decisions are made. Research shows the urgent need to develop principles and guidance to ensure ethical AI use in healthcare. This study aimed to explore healthcare professionals' perceptions of the ethical aspects of implementing an AI application to predict the mortality risk of patients in emergency departments. The analysis used an abductive qualitative content analysis based on the principles of medical ethics (autonomy, beneficence, non-maleficence, and justice), the principle of explicability, and the new principle of professional governance, that emerged from the analysis. In the analysis, two conflicts and/or considerations emerged tied to each ethical principle elucidating healthcare professionals' perceptions of the ethical aspects of implementing the AI application in emergency departments. The results were related to aspects of sharing information from the AI application, resources versus demands, providing equal care, using AI as a support system, trustworthiness to AI, AI-based knowledge, professional knowledge versus AI-based information, and conflict of interests in the healthcare system.

Keywords. AI applications, ethical aspects, healthcare professionals, qualitative study

1. Introduction and Methods

Artificial intelligence (AI) has shown promising potential for prediction and diagnostic classification in support of clinical decision-making [1]. One potentially effective use of predictive analytics is to support decisions on which patient would benefit the most from being admitted to the hospital wards and who could be safely discharged or displaced, resulting in optimized use of healthcare recourses and quality of health care. Research has shown the advantages of using predictive analytics in emergency departments [2], and to investigate this and potential ethical barriers to its use in practice, this study aimed to explore healthcare professionals' perceptions of the ethical aspects of implementing AI applications to predict mortality in emergency departments.

The study had a qualitative design [3] with an abductive approach using well-established ethical principles as framework for the deductive analysis [4, 5]. Individual interviews were conducted with 18 healthcare professionals and managers. The interviews were based on open-ended questions about implementing AI into healthcare

[1] Corresponding Author: Lena Petersson, E-mail: Lena.Petersson@hh.se.

and addressed the topic of ethical aspects of implementing the specific AI application for mortality prediction of emergency department patients. The study conformed to the Declaration of Helsinki requirements on research: information, consent, confidentiality, and safety of the participants. The analysis used an abductive qualitative content approach with pairing of meaning units to corresponding categories of the principles of medical ethics i.e autonomy, beneficence, non-maleficence, and justice [4] and the principle of explicability (addressed in the guidelines for trustworthily AI) [5]. Data that could not be mapped into the preexisting principles were given a new category.

2. Results, Discussion and Conclusions

The results describe ethical aspects of implementing AI applications for mortality prediction of emergency department patients. The findings align with the principles of medical ethics and the principle of explicability. In addition, a new principle of professional governance emerged and was suggested as a supplementary ethical principle in this context (see Table 1). This study provides insights from healthcare professionals' perspectives on the ethical aspects of implementing AI to predict mortality of emergency department patients. Our findings also demonstrate the need to address the six ethical principles to guide the implementation of AI applications in practice.

Table 1. Overview of the categories and subcategories describing ethical aspects of implementing AI-based applications to predict mortality in emergency departments

Categories/ Ethical principles	Subcategories/ conflicts/considerations
Autonomy	Conflicts between availability and sharing of AI-based information Considerations of using AI-based applications in relation to patients' self determination
Justice	Conflicts between demands and availability of resources in relation to implementing AI-based applications Considerations of using AI-based applications to systematically provide equitable healthcare
Beneficence and non-maleficence	Considerations of AI-based applications as a support system for physicians in their clinical practice Considerations of AI-based applications acting in the patient´s best interest
Explicability	Conflicts of AI technology's value and user-friendliness in clinical work Considerations of the trustworthiness of AI-based applications
Professional governance	Conflicts between AI-based information and physicians' experience-based knowledge AI-based conflicts of interest within and between healthcare organizations

References

[1] Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull. 2021;139:4-15.
[2] Boonstra A, Laven M. Influence of artificial intelligence on the work design of emergency department clinicians a systematic literature review. BMC health services research. 2022;22:669.
[3] Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse education today. 2017;56:29-34.
[4] Gillon R. Medical ethics: four principles plus attention to scope. Bmj. 1994;309:184.
[5] European C, Directorate-General for Communications Networks C, Technology. Ethics guidelines for trustworthy AI: Publications Office; 2019.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230235

Healthcare Leaders' Perceptions of the Usefulness of AI Applications in Clinical Work: A Qualitative Study

Lena PETERSSON[a,1], Petra SVEDBERG[a], Jens M NYGREN[a] and Ingrid LARSSON[a]
[a] *School of Health and Welfare, Halmstad University, Halmstad, Sweden*

Abstract. Artificial intelligence (AI) is often presented as a technology that changes healthcare and is useful in clinical work in disease prediction, diagnosis, treatment effectiveness, and precision health. This study aimed to explore healthcare leaders' perceptions of the usefulness of AI applications in clinical work. The study was based on qualitative content analysis. Individual interviews were conducted with 26 healthcare leaders. The usefulness of AI applications in clinical care was described in terms of expected benefits for 1) patients as supporting individualized self-management and person-centered information support tools 2) healthcare professionals in terms of providing decision-support in diagnostics, risk assessments, treatment recommendations, warning systems, and as a new colleague supporting the clinical work, and 3) organizations as providing patient safety and decision-support in prioritizing healthcare resources in organizing healthcare.

Keywords. AI-based applications, healthcare leaders, qualitative study, usefulness

1. Introduction

Artificial intelligence (AI) is often thought of as a technology that will transform healthcare through improvements in diagnostics, monitoring, access, advanced decision-making, and virtual consultations [1]. Research has anticipated that AI and precision medicine have the potential to revolutionize healthcare by identifying patients with special care needs through insights and learning from complex data [2]. However, empirical evidence for how to succeed with implementing AI in healthcare is scarce and not in proportion to the current needs in practice [3]. To address this, more knowledge is needed on the usefulness of AI applications from the perspectives of healthcare leaders and to further understand their roles as both gatekeepers and facilitators for successful implementation. This study, therefore, aimed to explore healthcare leaders' perceptions of the usefulness of AI applications in clinical work.

2. Methods

The study had an inductive qualitative content analysis approach [4]. Individual semi-structured interviews were conducted with 26 healthcare leaders who were in a position to potentially influence the implementation and use of AI systems in clinical practice in

[1] Corresponding Author: Lena Petersson, E-mail: Lena.Petersson@hh.se.

Sweden. The participants consisted of men (n=18) and women (n=8); professional roles were politicians (n=4), managers (n=9), and quality developers (n=13); work contexts were healthcare administration (n=15), primary care (n=6), psychiatry (n=3), and hospital care (n=2). In the analysis, meaning units were abstracted into codes, eight subcategories, and three categories [4]. The study was approved by the Swedish Ethical Review Authority (2020-06246) and conformed to the Declaration of Helsinki.

3. Results

Healthcare leaders' perceptions of the usefulness of AI applications in clinical work were that implementing AI applications would result in quality improvements beneficial for patients, healthcare professionals, and organizations (see Table 1).

Table 1. Overview of categories and subcategories describing the usefulness of AI applications in clinical work

Categories	Subcategories
Expected benefit for patients	AI as an individualized self-management tool
	AI as a person-centered information support tool
Expected benefit for healthcare professionals	AI as a decision support tool in diagnostics
	AI as a decision support tool in risk assessments and treatment recommendations
	AI as a warning system
	AI as a new colleague supporting the clinical work
Expected benefit for organizations	AI as a facilitator of patient safety
	AI as a decision support tool in prioritizing healthcare resources

4. Conclusion

This study highlighted healthcare leaders' perceptions of the usefulness of AI applications on different levels in healthcare, meeting the needs of patients, healthcare professionals, and organizations. The result adds knowledge about the perceived complexity of AI applications varied in different areas, although the usefulness was significant in different ways. AI research in diagnostic imaging is extensive [5], whereas for prediction of mental health outcomes is scarce [3]. Still, the leaders believed that AI could be useful in those areas where research and development remain.

References

[1] Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future of Health Care. Am J Med. 2019;132:795-801.
[2] Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision Medicine, AI, and the Future of Personalized Health Care. Clin Transl Sci. 2021;14:86-93.
[3] Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial Intelligence Applications in Health Care Practice: Scoping Review. Journal of medical Internet research. 2022;24:e40238.
[4] Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today. 2017;56:29-34.
[5] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31-8.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230236

Usability-Tests of Mechanical Ventilators: A Systematic Review

Julia S. VOLMERG [a,1], Beatrice COLDEWEY [a], Rainer RÖHRIG [a] and Myriam LIPPRANDT [a]
[a] *Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Germany*
ORCiD: Julia S. Volmerg: 0000-0003-4945-4604, Beatrice Coldewey: 0000-0001-9564-9467, Rainer Röhrig: 0000-0002-0032-5118, Myriam Lipprandt: 0000-0001-9371-0551

Abstract. The usability of ventilators is critical for patient safety. This systematic review shows the methods used in usability studies on ventilators, if those are similar in methodology. Furthermore, the usability tasks are compared to the requirements for manufactures during approval. Results show that the methodology and procedure of the studies are similar, but only cover part of the primary operating functions from their corresponding ISO Norm. Therefor optimisation of aspects of the study design, e.g., scope of tested scenarios, is possible.

Keywords Mechanical Ventilators, Usability-Test, User-Centered Design, Systematic Review, Patient Safety

1. Introduction

While technical progress increases complexity in medical devices like ventilators, multimorbidity and rising numbers of elderly people lead to more people in need of those machines. Compounding to this situation are unforeseen pandemics like COVID-19.

This leads to a rising number of people requiring ventilation, increasing the need for usability testing of ventilators. To avoid hazards for patients, manufacturers are bound by the Medical Device Regulation (MDR) to comply with the general safety and performance requirements. Therefore, a usability engineering process must be carried out on the primary operating functions [1]. The goal is to ensure that use errors either do not occur or do not result in harm.

The aim of this systematic review is to compare methodological approaches in usability tests of ventilators for similarities, and an unwittingly used standardization. The results are also compared to the requirements for the approval of medical devices.

2. Methods

PubMed and Embase were searched with a string consisting of "mechanical ventilation", "usability" and suitable synonyms, limited to the last 20 years. In accordance with the

[1] Corresponding Author: Julia S. Volmerg, E-mail: jvolmerg@ukaachen.de.

Prisma standard [2], relevant studies were filtered out by two independent researchers using the software Rayyan [3]. These studies were examined for various criteria such as study design, qualification and number of participants, usability test procedure, the individual tasks, and methods used to measure the outcomes. In order to assess the quality of the studies, an assessment was carried out using QualSys [4].

3. Results

Of the 5463 articles found, 12 relevant articles were identified. All studies conducted a comparative evaluation of usability in form of a summative test. The number of participants varied from 6 to 48, and from 2 to 13 tested devices. The process leading to sample size was documented in only four studies. Between 5 and 20 tasks were tested, which differed in their granularity. For comparability, these were mapped to the primary operating functions of the associated standards, e.g. ISO 80601-2-12 [5]. Regarding the usability outcomes, effectiveness (success rate), efficiency (time) were measured and satisfaction were determined in most of the studies by means of questionnaires. Six of the studies also measured the (perceived) workload.

4. Discussion

The review identifies differences as well as shortcomings in the study design. The results show that the methodology and the procedure of the studies are similar, but improvable. Furthermore, they are only partly comparable to the primary operating functions. This adds to the non-standardized way of usability testing, that could be improved by aligning the tasks at least the given categories in the norms. In some studies, the description of the method did not allow any reproducibility. There were no publications by manufacturers themselves. In order to improve the quality of usability studies of ventilators, more focus could be placed on the optimisation of certain aspects regarding the study design, e.g. the scope of tested scenarios with regard to the main operating functions, hazard-related usage scenarios or the systematic collection of usage problems. The limiting factor to this study is the use of different terms and synonyms in usability and medicine, leading to possible missed studies.

References

[1] ISO. ISO 62366-2:2015: Medical devices Part 2: Guidance on the application of usability engineering to medical devices 04-2016. 1st ed. Geneva: ISO International Organization for Standardization.
[2] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.
[3] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5:210. doi:10.1186/s13643-016-0384-4.
[4] Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Edmonton, Alta.: Alberta Heritage Foundation for Medical Research; 2004.
[5] ISO 80601-2-12:2020-07: Particular requirements for basic safety and essential performance of critical care ventilators.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230237

Towards Implementation of a Home-Based Phantom Limb Pain Treatment Facilitated by Textile-Electrode System - A Case Study

Anna BJÖRKQUIST[a], Li GUO[a,1], Morten B. KRISTOFFERSEN[b], Maria M. NOVOA[b], Max ORTIZ-CATALAN[b] and Leif SANDSJÖ[c]
[a] *Swedish School of Textiles, University of Borås, Borås, Sweden*
[b] *Center for Bionics and Pain Research, Mölndal, Sweden*
[c] *Dept. of Work Life and Social Welfare, University of Borås, Borås, Sweden*
ORCiD ID: Anna BJÖRKQUIST https://orcid.org/0000-0001-7632-8262 ,
Li GUO https://orcid.org/0000-0002-1949-3365,
Maria M. NOVOA https://orcid.org/0000-0002-6001-985X
Morten B. KRISTOFFERSEN https://orcid.org/0000-0002-3901-2856,
Max ORTIZ-CATALAN https://orcid.org/0000-0002-6084-3865,
Leif SANDSJÖ https://orcid.org/0000-0001-5338-769X

Abstract. This case study reports the use of a new textile-electrode system for self-administered Phantom Motor Execution (PME) treatment at home in one patient with Phantom Limb Pain (PLP). In follow-up interviews, the patient reported reduced pain, increased mobility, and improved mental health, and aspects such as motivation, usability, support, and treatment outcome, could be recognized from an earlier study as crucial for successful implementation and adoption of the home-based long-term treatment. The findings are of interest to developers, providers, users, and researchers planning home-based clinical studies and/or scenarios based on technology-assisted treatment.

Keywords. Phantom limb pain, Textile electrodes, Home-based, Self-administered

1. Introduction

Phantom Limb Pain (PLP) can be treated with Phantom Motor Execution (PME), which involves controlling virtual limbs based on myoelectric pattern recognition [1]. However, the conventional Ag/AgCl electrodes used to record the myoelectric signal may not be ideal for home-based treatment compliance. A pilot study showed that a textile-electrode system is better suited for PLP treatment at home [2]. This case study investigates the effectiveness of a textile-electrode system for home-based, self-administered, and long-term PME treatment, using previous user experiences from a study on neuromuscular conditions as a framework [3].

[1] Corresponding Author: Li Guo, University of Borås, Allégatan 1, 50190, Borås, Sweden. E-mail: li.guo@hb.se.

2. Methods

The participant, John (pseudonym), was provided with a PME treatment system that included a myoelectric pattern recognition system and a custom-made textile-electrode system to use at home for 2 periods of 12 weeks each with and without a prescribed training recommendation. He had tried several treatments for his PLP with limited success but had previously used the PME treatment at a clinic with good results. Interviews were conducted after each of the two training periods. The interviews were analyzed using a deductive approach focusing on *motivation*, *usability*, *support*, and *outcome* from using the textile electrode system enabled PME treatment at home.

3. Results

John's responses to the two interviews are summarized according to findings in [3]. *Motivation*: John's primary motivation for participating in the PME treatment was to reduce his pain, but he also found motivation from being part of a research project and the enjoyment of training. *Usability*: John found the entire PME system, especially the textrode-band, to be technically superior and easier to use than the conventional Ag/AgCl electrodes which he found both uncomfortable and time-consuming to use. *Support*: While John appreciated to have a physiotherapist "around" to place the electrodes during his previous treatment, he now felt empowered by being able to train on his own. *Outcomes*: John experienced a significant reduction in pain intensity and incidence, which had led to increased mobility, improved mental health and quality of life.

4. Discussion and Conclusion

This case study explores home-based, self-administered, and long-term PLP treatment with a textile-electrode system. The participant's experiences provide valuable insights on how to improve PLP treatment at home, including the importance of usability, motivation, and empowerment of patients. These insights may be valuable to developers, technology providers, researchers, and users of assistive technology on what to consider when introducing home-based treatment concepts. We believe the results to be transferable to other treatment initiatives and areas that may benefit from home-based and self-administered scenarios, such as post-stroke rehabilitation and fibromyalgia.

References

[1] Ortiz-Catalan M. The Stochastic Entanglement and Phantom Motor Execution Hypotheses: A Theoretical Framework for the Origin and Treatment of Phantom Limb Pain. Front Neurol. 2018 Sep 6;9:748. doi: 10.3389/fneur.2018.00748. PMID: 30237784; PMCID: PMC6135916.

[2] Lendaro E, Guo L, Novoa MJM, et al. (2019). Seamless Integrated Textrode-Band for Real-time Lower Limb Movements Classification to Facilitate Self-Administrated Phantom Limb Pain Treatment. Conference proceedings: IEEE Engineering in Medicine and Biology Society. Conference. 2019. 1753-1756. 10.1109/EMBC.2019.8856979.

[3] Jonasson LL, Sörbo A, Ertzgaard P, Sandsjö L. Patients' Experiences of Self-Administered Electrotherapy for Spasticity in Stroke and Cerebral Palsy: A Qualitative Study. J Rehabil Med. 2022 Feb 14;54:jrm00263. doi: 10.2340/jrm.v53.1131. PMID: 34935050; PMCID: PMC8892303.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230238

Usability Assessment as a Participatory Method for Implementing Technical Innovations in Nursing Care

Jörn KRÜCKEBERG[a,1], Isabel OTTMANN[a], Daniel Peter BEUME[b], Ronny KLAWUNN[c] and Nicole HECHTEL[a]
[a] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hannover Medical School, Germany*
[b] *Hochschule Hannover - University of Applied Sciences and Arts, Germany*
[c] *Institute for Epidemiology, Social Medicine and Health Systems Research, Hannover Medical School, Germany*

Abstract. For the introduction of technical nursing care innovations, a usability assessment survey is conducted by nursing staff. The questionnaire is used before and after the introduction of technical products. This poster contribution shows the latest comparison of pre- and post-surveys on selected products.

Keywords. Nursing technology, participation management, nursing innovation, usability assessment

1. Introduction

The research project Centre of Implementing Nursing Care Innovations (PPZ Hannover) aims to integrate innovative technologies to support nursing staff and to improve patient care. To ensure the participation of nurses' perspective in implementation decisions, a participatory implementation concept was designed [1]. One part of the concept are workshops, in which the innovative technologies are discussed with nursing staff as the primary user-group.

2. Method

As an important decision instrument, a modified questionnaire for assessing the usability of medical devices [2] is used during the workshops with the nursing staff to assess the expected usability and feasibility. About three months after the introduction, the questionnaire is conducted again to reveal the experiences made with the innovation by the nursing staff. The questionnaire contains general questions that can be used for all technical products and specific questions that need to be selected and customized according to the products´ specifics. The questions can be answered on a 5-point-Likert

[1] Corresponding author: Jörn Krückeberg, Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig - Institute of Technology and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover; E-mail: Joern.Krueckeberg@plri.de.

Scale ranging from statement disagreement to full agreement. As result, a value named the “weighted usability” is summarized that distinguishes between "not suitable", “suitable to a limited extent" and "very suitable".

3. Results

The following table shows the usability values of three products in a pre/post comparison [Table 1]. The indicated number presents a single nurse that assessed the usability of the product. The mattress system is used to mobilize patients for decubitus prophylaxis, the robotic cat is used for activating and interacting with patients, and the sound cushion produces relaxing music to improve patients' state of restlessness.

Table 1. The table shows results of the usability assessment survey before and after the implementation of three technical products on the project ward at Medical School Hannover. Eight to sixteen nurses participated in the surveys.

		Product					
		Mattress System (pre) n=8	Mattress System (post) =12	Robotic cat (pre) n=15	Robotic cat (post) n=11	Sound cushion (pre) n=16	Sound cushion (post) n=9
Results	Not suitable	0	1	2	2	0	0
	Suitable to a limited extent	0	2	6	2	0	0
	Very suitable	8	9	7	7	16	9

The mattress system as well as the sound cushion are rated as "very suitable" in both the pre- and post-surveys. Although the response between the pre- and post-survey differs, both products are expected to be very suitable in everyday nursing care and are still rated as very suitable several months after the implementation. The interactive robot cat receives mixed feedback with regard to its suitability in everyday nursing care.

4. Discussion

The results of the pre-survey showed a fairly clear tendency to introduce the products and the mostly positive results of the post-survey give a consistent picture in the pre/post comparison, proving the accuracy of the implementation decision. Therefore, the questionnaire can be seen as a supporting element in the process of a participatory approach for introducing technical products.

References

[1] Klawunn R., Dierks M. D., Krückeberg J., & Hechtel N. (2021). Partizipation von Pflegefachpersonen in Technikauswahl und -einführung. Public Health Forum 29(3): 260–264. https://doi.org/10.1515/pubhef-2021-0065

[2] L. Müller, C. Backhaus: Entwicklung eines Fragebogens zur Bewertung der Gebrauchstauglichkeit von Medizinprodukten im Beschaffungsprozess. In: GfA Dortmund (Ed.): Frühjahrskongress Dresden 2019.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230239

Physician-Reported Experience and Usability of the MyPal Platform: A Palliative Care Digital Health Intervention

Panos BONOTIS[a,1] and Pantelis NATSIAVAS[a]
[a]*Institute of Applied Biosciences, Centre for Research and Development Hellas, Thessaloniki, Greece*
ORCiD ID: Panos Bonotis https://orcid.org/0000-0002-9317-547X
Pantelis Natsiavas https://orcid.org/0000-0002-4061-9815

Abstract. There is a lack of research focusing on the physician-end, their experiences, and their perception of usability with an eHealth intervention. The aim of this study was to evaluate physician satisfaction, and perception of usability following the use of the MyPal platform, a digital health intervention to foster palliative care for hematological cancer patients. Participants were healthcare professionals active in the project's multinational randomized clinical trial evaluating the impact of the MyPal platform. A post-study electronic questionnaire was administered comprised of; 2 standardized questionnaires (PSSUQ, UEQ) and a feature satisfaction questionnaire, and an open ended question. All questionnaire scores were relatively high and the platform was more than marginally accepted by all participants.

Keywords. eHealth, palliative care, physician satisfaction, usability

1. Introduction and Background

There is a clear demand based on the literature for improvements in the designed digital health interventions for palliative cancer care on the basis of all stakeholder needs [1]. A recent review of digital health interventions across multiple diseases found that usability from the physicians side was assessed in only 33% of studies [2].

MyPal[2] is a collaborative H2020 research project aiming to use eHealth technologies in order to support palliative care for cancer patients and healthcare professionals (HCPs), via the adoption of the electronic Patient Reported Outcomes (ePRO) paradigm. The MyPal platform includes an HCP web application, with various modules to support the HCPs, where mainly ePRO data collected from patients are presented in an interactive dashboard via data-intensive visual analytics [3]. The presented study's primary objective was to evaluate HCPs perception of usability and satisfaction with the use of the HCP web app, and overall experience of using the system during the RCT in parallel

[1] Corresponding Author: Panos Bonotis, eHealth Lab, INAB-CERTH, 6th Km. Charilaou-Thermi Road, Thermi, Thessaloniki, Greece; E-mail: pbonotis@certh.gr.

[2] The MyPal project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825872, https://mypal-project.eu/

with their routine obligations via standardized quantitative questionnaires after sustained use of a digital health intervention using this web application.

2. Methods and Results

The participants of this study were 14 HCPs that participated in the MyPal Randomized Clinical Trial (RCT). They used the MyPal HCP web platform for approximately a 12 month period, during which the HCPs had various obligatory and optional actions to perform via the MyPal HCP web platform[3]. After the whole MyPal study was completed, a study follow-up questionnaire was administered electronically (via the MyPal platform) to the HCPs. It consisted of: (a) the Post Study System Usability Questionnaire (PSSUQ) that scored 2.32 (sd: 1.15). The highest score was noted in the system usefulness subdomain (2.01, sd: 0.95). The platform was rated at 2.64 (sd: 1.17) in regard to interface quality and at 2.42 (sd: 1.17) for information quality. (b) The short version of the User Experience Questionnaire (UEQ) which scored 0.661 (sd: 1.25). The platform was rated at 0.679 (sd: 1.35) in the hedonic category which relates to the design, the aesthetical perceptions and 0.643 (sd: 1.15) in the pragmatic category. (c) A feature-specific satisfaction questionnaire to collect users' perceived satisfaction for each feature of the platform with scores ranging from 1 to 7, lower being better. The features means were calculated with relatively high scores and a mean across all features at 2.04 (sd: 1.27). And finally, an optional open-ended question was answered by 2 HCPs reporting on technical problems and the need for a better management system of the users' entries.

3. Discussion and Conclusions

This study offers insights into physician satisfaction, preferences, and perceptions of the MyPal HCP web platform. Based on the participants' feedback, the MyPal HCP web platform is accepted in terms of usability, and noticeably the system's usefulness is highlighted by all participants responses. The satisfaction levels of the individual features of the platform are also acceptable. Although in terms of usability the web application is relatively well accepted, it should be noted that big fixes were sometimes necessary, and this could significantly affect the user experience evaluation. Moreover, based on the answers provided to the open-ended questions some extra features would also be useful.

References

[1] Karamanidou C, Natsiavas P, Koumakis L, Marias K, Schera F, Schäfer M, et al. Electronic Patient-Reported Outcome-Based Interventions for Palliative Cancer Care: A Systematic and Mapping Review. JCO Clin Cancer Inform. 2020 Jul 22;4(4):647–56.

[2] Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: A systematic review of the literature. JMIR Med Inform. 2016;4(1).

[3] Koumakis L, Schera F, Parker H, Bonotis P, Chatzimina M, Argyropaidas P, et al. Fostering Palliative Care Through Digital Intervention: A Platform for Adult Patients With Hematologic Malignancies. Front Digit Health. 2021 Dec 17;0:186.

[3] MyPal platform demo video: https://youtu.be/K32nGL2R7sk

This page intentionally left blank

Section 6

Knowledge and Information Representation and Modeling

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230241

Machine Learning for Medical Data Integration

Armin MÜLLER [a,1], Lara-Sophie CHRISTMANN [b], Severin KOHLER [b], Roland EILS [b] and Fabian PRASSER [a]

[a] *Center of Health Data Science, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany*
[b] *Digital Health Center, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany*

ORCiD ID: Armin Müller https://orcid.org/0000-0003-3566-8687,
Lara-Sophie Christmann https://orcid.org/0000-0001-6453-8633,
Severin Kohler https://orcid.org/0000-0002-7718-6187,
Roland Eils https://orcid.org/0000-0002-0034-4036,
Fabian Prasser https://orcid.org/0000-0003-3172-3095

Abstract. Making health data available for secondary use enables innovative data-driven medical research. Since modern machine learning (ML) methods and precision medicine require extensive amounts of data covering most of the standard and edge cases, it is essential to initially acquire large datasets. This can typically only be achieved by integrating different datasets from various sources and sharing data across sites. To obtain a unified dataset from heterogeneous sources, standard representations and Common Data Models (CDM) are needed. The process of mapping data into these standardized representations is usually very tedious and requires many manual configuration and refinement steps. A potential way to reduce these efforts is to use ML methods not only for data analysis, but also for the integration of health data on the syntactic, structural, and semantic level. However, research on ML-based medical data integration is still in its infancy. In this article, we describe the current state of the literature and present selected methods that appear to have a particularly high potential to improve medical data integration. Moreover, we discuss open issues and possible future research directions.

Keywords. medical data integration, common data models, machine learning

1. Introduction

Insights generated through data-driven medical research methods based on the secondary use of health data have the potential to make future medicine more predictive, preventive and personalized [1]. This can improve cost efficiency and foster adoption to demographic change. To make health data from a diverse set of sources available for research and enable real-world evidence generation, they need to be integrated into common data models that facilitate comparability. Well-known models in the health field in-

[1]Corresponding Author: Armin Müller, Center of Health Data Science, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany; E-mail: armin.mueller@charite.de

clude the OMOP Common Data Model [2], HL7 Fast Healthcare Interoperability Resources [3] and openEHR [4]. When data was not collected in accordance to such models at source – which is common to date – it needs to be harmonized and transformed.

Due to the large number of autonomous information systems within typical health IT infrastructures, data is usually heterogeneous along three axes: (1) syntax (e.g. regarding the meaning of symbols), (2) structure (e.g. regarding the organization of properties of health-related data entities) and (3) semantics (e.g. regarding terminologies as well as codes and their meaning).

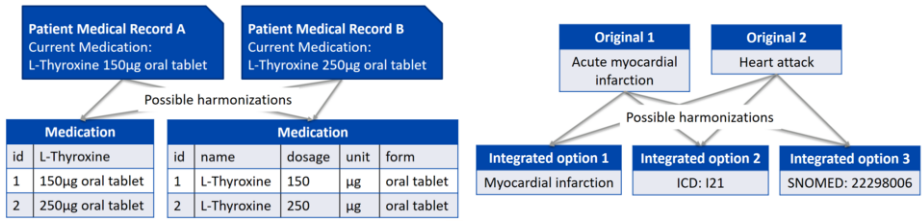


Figure 1. Example for structural heterogeneity (left) and semantic heterogeneity (right).

An example for syntactical heterogeneity are the different possible encodings of the character "μ" in the drug prescription "L-Thyroxine 150μg oral tablet", which can be encoded as "\u03BC" in Unicode, as "230" in ASCII, and as "μ" in HTML. Figure 1 illustrates a simple example of structural and semantic heterogeneity. On the left, the same information about the administration of a drug is structured in different ways. The right side demonstrates that multiple terms can refer to the same concept.

To date, harmonization is mostly achieved by using manually specified rules and algorithms in so-called Extract-Transform-Load (ETL) processes, supported by tools such as Pentaho Data Integration [5] or Talend Open Studio [6]. While innovative approaches, e.g. based on declarative specifications of target representations [7], can help to reduce some efforts, this process is usually still very time and resource consuming [8].

ML and Artificial Intelligence (AI) technologies are one of the core tools of data-driven medical research. The general idea is that instead of providing computers with rules to follow, they extract knowledge and discover rules themselves from training data provided [9]. For these algorithms and models to produce reasonable results, they rely heavily on large datasets, clearly demonstrating the need for medical data integration. With their predictive and generative capabilities, ML methods can also potentially be a powerful tool for data integration itself. In recent years, machine learning has already been applied very successfully to various knowledge extraction and standardization tasks. One important field is natural language processing, where ML has been very successful at understanding the structure of clinical documents [10] and extracting medical concepts from clinical free-text reports [11].

2. Objective

The aim of this paper is to investigate the potential of ML for medical data integration tasks and to provide a concise overview of the current state of the field. We focus on methods suited for integrating structured health data into standardized data models. More specifically, this paper presents examples of methods that have been suggested for ML-based harmonization of data on the syntactic, structural and semantic level. We fur-

ther discuss their potential value for medical data integration tasks and highlight limitations as well as open research questions.

3. Method

Early papers from the data integration community have suggested that ML could be used to automate or support many of the tasks needed to harmonize and integrate structured data. Starting from a seminal paper by Dong and Rekatsinas, which, to our knowledge, is the first – and one of the only ones so far – to address the potential of ML for data integration tasks [12], we reviewed the state of the literature and present selected highlights.

We screened all of the papers citing the work by Dong and Rekatsinas [12] ($n = 117$ in December 2022). We excluded nine non-English papers ($n = 108$) and then selected all papers whose titles indicate a possible focus on data integration ($n = 78$). From this, we selected all papers with abstracts suggesting that they address structured tabular data or health data. We also excluded reviews as they only covered narrow aspects of the general topic relevant to our work. This resulted in $n = 22$ papers of which 14 addressed the topic of entity resolution (sometimes also called entity matching), which is an important aspect of semantic integration that seems to have received quite some scientific attention.

In the following section, we present in more detail selected papers from the body of literature identified in the described search process as well as selected papers discovered during a preparatory exploration of the field.

4. Results

4.1. Syntactic Heterogeneity

While it is already difficult to automatically detect the character set with which files are encoded [17], determining the composition and types of data items is even more challenging. One aspect that is particularly relevant for structured data integration, focuses on extracting the orientation and sub-components of tables, which is an important first step in any transformation process. Recently, Habibi et al. proposed a deep learning method for classifying table orientation, achieving an F_1-score of 76% [18]. Other relevant systems include MIT's Sherlock, which can detect data types (e.g. dates) in structured data with an F_1-score of 89% using deep neural networks [19].

4.2. Structural Heterogeneity

Just recently, Sahay et al. studied self-organizing maps combined with a priori knowledge about the target structure to conquer structural heterogeneity, achieving an F_1-score of 71% [20]. Anderson inspected column embeddings as input into a bidirectional LSTM model to label columns and tilt tables [13]. Both approaches are applicable to target data models with a specific pre-defined structure, such as the OMOP CDM. Toutanova et al. used neural nets to extract facts in the form of (subject, predicate, object)-triples, outperforming prior work in precision [21]. This is suited for mapping to generic models based on fact-tables, such as Informatics for Integrating Biology & the Bedside (i2b2) [22].

4.3. Semantic Heterogeneity

To tackle semantic heterogeneity, Kate used support vector machines to map terms from clinical narratives to SNOMED CT codes [23], achieving an F1-score of 88%. Wang et al. proposed using contrastive representation learning to facilitate multiple data integra-

tion tasks including entity and column matching [16]. Parr et al. focused on lab values and LOINC codes using logistic regression and a random forest multiclass classifier [24], achieving an F1-score of up to 62%, while Mirzaei et al. compared a logistic regression classifier, a random forest classifier and a fully connected neural network classifier to standardize variables within and across datasets [14]. Recently, Zhang et al. have approached the problem of detecting the semantic type of data items using a deep neural network for single column predictions. The results can then be forwarded into a multi-layer structured prediction model that outputs the final classification per column [25].

5. Discussion

In this paper we presented a concise overview of selected ML-based methods supporting core steps of medical data integration. In theory, a combination of such individual methods could be used to develop end-to-end ML-based data integration processes. In practice, however, several challenges have to be overcome to make this vision a reality.

First, most of the current solutions provide a performance that is not sufficient to reliably support data integration processes without a lot of manual intervention (cf. the F_1-scores presented in Section 4). Hence, further research on improved methods and human-in-the-loop approaches is needed. For example, Graph Neural Networks seem promising to improve the accuracy of structural data integration steps [15]. Since health data is inherently different from many other data domains - e.g., due to it being longitudinal and sometimes of low quality - the applicability of methods developed for non-health data remains to be evaluated.

Second, in addition to novel methods, also more comprehensive training and test sets are needed. Given the amount of work that has already went into medical data integration on a global scale, we are confident that large enough sets of matching original and harmonized data could be created. Nonetheless, they would also need to be curated and sharing them will likely pose privacy risks. An additional option would be to integrate knowledge about standardized data representations, e.g. from openEHR, FHIR or the OMOP CDM, into the ML models.

Finally, approaches are needed to introduce ML-based methods incrementally and in synergy with more traditional processes, e.g. as atomic operators in common data integration platforms. Considering the vast amount of heterogeneity in health data, we believe that this would be a significant step for advancing medical research.

In future work, it would be interesting to benchmark combinations of ML-based integration approaches along multiple axes and to compare their performance in terms of their integration accuracy and scalability. Furthermore, the work presented in this paper only provides a first overview of the topic. In the future, we aim to build upon this work by performing a more in-depth structured review on ML-based medical data integration.

References

[1] Hood L, Balling R, Auffray C. Revolutionizing Medicine in the 21st Century Through Systems Approaches. Biotechnol J. 2012 Aug;7(8):992-1001. doi: 10.1002/biot.201100306. Epub 2012 Jul 20.

[2] Voss EA, Makadia R, Matcho A, et al. Feasibility and Utility of Applications of the Common Data Model to Multiple, Disparate Observational Health Databases. J Am Med Inform Assoc. 2015 May;22(3):553-64. doi: 10.1093/jamia/ocu023. Epub 2015 Feb 10. PMID: 25670757; PMCID: PMC4457111.

[3] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful Approach to Healthcare Information Exchange. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013 Jun; pp. 326-331, doi: 10.1109/CBMS.2013.6627810.
[4] Kalra D et al. The openEHR Foundation. Stud Health Tech Inform. 2005;115:153-73. PMID: 16160223.
[5] Casters M, Bouman R, van Dongen J. Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. Indianapolis, IN: Wiley, Inc; 2010. 720p. ISBN: 9780470635179.
[6] Talend Open Studio: Open-source ETL and Free Data Integration. Talend. Available at: https://www.talend.com/products/talend-open-studio/ (Accessed: December 12, 2022).
[7] Spengler H, Lang C, Mahapatra T, Gatz I, Kuhn K, Prasser F. Enabling Agile Clinical and Translational Data Warehousing: Platform Development and Evaluation. JMIR Med Inform 2020;8(7):e15918. doi: 10.2196/15918.
[8] Kimball R, Ross M. The Data Warehouse Toolkit: The definitive guide to dimensional modeling. Indianapolis, IN: John Wiley & Sons, Inc.; 2013. 608p. ISBN: 9781118530801.
[9] Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016 Sep 29;375(13):1216-9. doi: 10.1056/NEJMp1606181. PMID: 27682033.
[10] Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications. J Am Med Inform Assoc. 2010 Sep-Oct;17(5):507-13. doi: 10.1136/jamia.2009.001560. PMID: 20819853.
[11] Kholghi M, Sitbon L, Zuccon G, Nguyen A. Active Learning: A Step Towards Automating Medical Concept Extraction. J Am Med Inform Assoc. 2016 Mar;23(2):289-96. doi: 10.1093/jamia/ocv069.
[12] Dong XL, Rekatsinas T. Data Integration and Machine Learning: A Natural Synergy. Proc. 2018 Int. Conf. Management of Data. 2018 May. pp. 1645–1650. doi: 10.1145/3183713.3197387.
[13] Anderson MR. It's Data All the Way Down: Exploring the Relationship Between Machine Learning and Data Management [dissertation]. [Michigan]: University of Michigan; 2019. 124p.
[14] Mirzaei A, Aslani P, Schneider CR. Healthcare Data Integration Using Machine Learning: A Case Study Evaluation with Health Information-seeking Behavior Databases. Res Social Adm Pharm. 2022 Dec;18(12):4144-4149. doi: 10.1016/j.sapharm.2022.08.001. Epub 2022 Aug 8. PMID: 35965198.
[15] Krivosheev E, Atzeni M, Mirylenka K, Scotton P, Casati F. Siamese Graph Neural Networks for Data Integration. eprint arXiv. 2020 Jan. doi: 10.48550/arXiv.2001.06543.
[16] Wang R, Li Y, Wang J. Sudowoodo: Contrastive Self-supervised Learning for Multi-purpose Data Integration and Preparation. arXiv preprint. 2022 Oct. doi: 10.48550/arXiv.2207.04122
[17] Kikui GI. Identifying, the Coding System and Language, of On-line Documents on the Internet. Proc. 16th Conf. Comp. Linguistics. 1996 Aug. pp. 652–657. doi: 10.3115/993268.993282.
[18] Habibi M, Starlinger J, Leser U. DeepTable: A Permutation Invariant Neural Network for Table Orientation Classification. Data Mining and Knowledge Discovery. 2020 Sep;34(6):1963-83. doi: 10.1007/s10618-020-00711-x.
[19] Hulsebos M, Hu K, Bakker M, et al. Sherlock: A Deep Learning Approach to Semantic Data Type Detection. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2019 Jul. pp. 1500-1508. doi: 10.1145/3292500.3330993.
[20] Sahay T, Mehta A, Jadon S. Schema Matching Using Machine Learning. 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN). 2020 Apr. doi: 10.1109/SPIN48934.2020.9071272.
[21] Toutanova K, Chen D, Pantel P, Poon H, Choudhury P, Gamon M. Representing Text for Joint Embedding of Text and Knowledge Bases. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015 Sep. doi: 10.18653/v1/D15-1174.
[22] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010 Mar-Apr;17(2):124-30. doi: 10.1136/jamia.2009.000893. PMID: 20190053.
[23] Kate RJ. Towards Converting Clinical Phrases into SNOMED CT Expressions. Biomed Inform Insights. 2013 Jun 24;6(Suppl 1):29-37. doi: 10.4137/BII.S11645. PMID: 23847425; PMCID: PMC3702194.
[24] Parr SK, Shotwell MS, Jeffery AD, Lasko TA, Matheny ME. Automated Mapping of Laboratory Tests to LOINC Codes Using Noisy Labels in a National Electronic Health Record System Database. J Am Med Inform Assoc. 2018 Oct 1;25(10):1292-1300. doi: 10.1093/jamia/ocy110. PMID: 30137378.
[25] Zhang D, Hulsebos M, Suhara Y, Demiralp Ç, Li J, Tan W-C. Sato: Contextual Semantic Type Detection in Tables. Proceedings of the VLDB Endowment. 2020 Jun;13(12):1835–1848. doi: 10.14778/3407790.3407793.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230242

Definition, Composition, and Harmonization of Core Datasets Within the German Center for Lung Research

Mark R. STÖHR [a,1], Andreas GÜNTHER [a] and Raphael W. MAJEED [a,b]
[a] *UGMLC, German Center for Lung Research (DZL), Justus-Liebig-University, Giessen, Germany*
[b] *Institute of Medical Informatics, Medical Faculty of RWTH Aachen, Aachen, Germany*

Abstract. Core datasets are the composition of essential data items for a certain research scope. As they state commonalities between heterogeneous data collections, they serve as a basis for cross-site and cross-disease research. Therefore, researchers at the national and international levels have addressed the problem of missing core datasets. The German Center for Lung Research (DZL) comprises five sites and eight disease areas and aims to gain further scientific knowledge by continuously promoting collaborations. In this study, we elaborated a methodology for defining core datasets in the field of lung health science. Additionally, through support of domain experts, we have utilized our method and compiled core datasets for each DZL disease area and a general core dataset for lung research. All included data items were annotated with metadata and where possible they were assigned references to international classification systems. Our findings will support future scientific collaborations and meaningful data collections.

Keywords. Data collection, datasets as topic, quality indicators, respiratory system, controlled vocabulary.

1. Introduction

One fundamental step towards successful cross-site and cross-disease research is the agreement on common fields of interest. These fields need to be identified and specified by naming and describing their essential data items.

The German Center for Lung Research (DZL) aims for high-impact translational research involving five participating sites as well as associated partners. Over 300 principal investigators contribute to eight disease areas: asthma and allergy, acute respiratory distress syndrome, cystic fibrosis, chronic obstructive pulmonary disease, diffuse parenchymal lung disease, end stage lung disease, lung cancer, and pulmonary hypertension. For advanced data science and collaborations, a central access point for data analysis and feasibility queries was established in 2016. Data from various heterogeneous sources is harmonized and integrated in a central data warehouse, currently containing over 60 individual data sources. Yet, the overlap of common parameters is low. In consequence, the DZL data warehouse does not sufficiently support cross-domain data analysis.

To achieve meaningful data for cross-site and cross-disease research, a common approach is to concentrate on parameters that promise to have the highest impact in terms

[1] Corresponding Author: M.R. Stöhr, E-mail: mark.stoehr@innere.med.uni-giessen.de.

of overarching availability and research relevance. We refer to them as "core parameters" that are collected and structured in "core datasets". Core datasets – by definition – contribute to the fulfilling of the FAIR principles [1], especially the reusability aspect of meeting domain-relevant community standards. The third FAIR principle "interoperability" can be achieved by referring to international terminologies like the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT).

Several national and international projects defined core datasets: the German Medical Informatics Initiative (MII) [2] has defined six modules for their core dataset, the German Interdisciplinary Association for Intensive Care and Emergency Medicine (DIVI) defined a core data set for emergency hospitalization [3], and, on an international scale, the BrainIT group gathered 25 experts from nine countries to compose a core dataset for patients with traumatic brain injury [4]. None of the core datasets developed so far meet the requirements for profound lung research.

Goal of this study is to compose nine core datasets, one for each of the eight DZL disease areas, and one disease area overarching DZL core dataset. Those core datasets need to be well annotated, e.g., by referring to standard terminologies.

2. Methods

2.1. Defining Criteria for Elements of Core Datasets

The DZL central data management proposed a definition on how to identify core parameters for lung health research. After revision by the DZL Board of Directors, the definition served as guideline for the following compositions of core datasets.

2.2. Compilation of core datasets for the DZL disease areas

Disease area coordinators were asked to name at least two domain experts per disease area to oversee defining the respective core datasets. The domain experts had to compose parameter lists fulfilling the criteria for core dataset elements. It was not specified how to find the parameters. Nevertheless, we proposed the methodology to take a large and well-established study as basis and filter the parameters that fulfill the requirements. The resulting parameter lists were discussed within each disease area for consensus.

In a second step, the lists from all eight disease areas were harmonized in uniform Excel sheets. List items were classified into three levels of semantic depth: "category", "parameter", or "characteristic" (e.g., "Biometric Data", "Gender", and "Female"). Within several feedback loops involving the domain experts, entries were complemented by generalization (e.g., "Gender" → "Biometric Data") and specification (e.g. "Specimen" → "Blood"). For unambiguity, we added columns for "data type" as well as a link to the entry in our metadata repository CoMetaR [5] if available.

2.3. Compilation of a Core Dataset for the DZL

The entries of all disease area core datasets were compared and overlapping entries identified. The DZL core dataset was then defined as all parameters covered by at least two disease areas. To assess the feasibility of future collection of these parameters from all disease areas, regardless of whether they defined it for their own core dataset, a

meeting with domain experts from all disease areas was called. All attendees were asked to which DZL core parameters their domain can contribute.

2.4. Integration of Core Dataset in the DZL Metadata Repository

All parameters from all core datasets were added to the DZL metadata repository CoMetaR if not yet included. Besides attributes like label, description, and data type, whenever applicable, a reference to international classifications like SNOMET-CT, ICD-10, and LOINC was added. We recorded how many parameters had to be added to CoMetaR and to what extent core datasets were covered by international classifications.

3. Results

3.1. Core Dataset Definition

The DZL leadership agreed on the following definition of a core dataset: "A disease area-specific dataset definition is required to enable cross-dataset evaluations within the disease area as well as across disease areas and to ensure good data quality. The dataset definition MUST contain all information required for a reliable diagnosis. These are parameters and criteria that are essential for correct phenotyping according to current guidelines. The dataset definition SHOULD include all relevant information on symptom burden, prognosis, quality of life (e.g. EQ5D), inclusion criteria as well as longitudinal course and extent of treatment."

3.2. Core Dataset Size and Terminology Coverage

We composed one DZL and eight plus one core datasets for all disease areas. The additional core dataset was created during the composition within the ARDS disease area. The disease "pneumonia", included in the ARDS disease area, turned out to have significantly distinct parameters. Table 1 shows the sizes of all core datasets, to what extent they were already included in the DZL metadata catalogue, and how well the core datasets are covered by international classifications. The number of parameters ranges from 16 to 159, the ratio of new parameters per core dataset ranges from three percent to 65 percent, and the coverage of international classifications ranges from 48 percent to 100 percent. The terminologies used for standardized code annotations were SNOMED-CT, International Classification of Diseases (ICD) German modification, Logical Observation Identifiers Names and Codes (LOINC), Anatomical Therapeutic Chemical (ATC) Classification System, and Operation and Procedure (OPS).

For quantitative view on the DZL core dataset, in the following paragraph, ARDS parameter coverage counts as true if both the ARDS and pneumonia core dataset cover it. Only two parameters were seen as core parameters by all eight disease areas, three by seven, two by six, nine by five, four by four, six by three, and 15 by two disease areas. The feasibility meeting resulted in nine parameters that are available for eight disease areas, nine parameters by seven disease areas, five by six, four by five, two by four, none by three, and eleven parameters by two disease areas.

The DZL core dataset is an intersection of the 9 disease area specific core datasets. The parameter "comorbidities" is an exception, as it is essential for all disease areas, but

the process of generalization and specification of core parameters yielded in different lists of comorbidities due to different research scope of the disease areas. Other than parameters like “smoking status” with definite characteristics, comorbidities are not immediately comparable across datasets.

Table 1. Shown are the core dataset domains, the number of parameters, the ratio of newly added parameters and the ratio of parameters that are listed in international classifications. DZL stands for German Center for Lung Research, AA for Asthma and Allergy, ARDS for Acute Respiratory Distress Syndrome, CF for Cystic Fibrosis, COPD for Chronic Obstructive Pulmonary Disease, DPLD for Diffuse Parenchymal Lung Disease, ELD for End stage Lung Disease, LC for Lung Cancer, and PH for Pulmonary Hypertension.

Domain	No. of Parameters	No. of New Parameters	No. of Parameters with intern. Codes
DZL	40	12 (30%)	34 (85%)
AA	51	30 (59%)	47 (92%)
ARDS	159	103 (65%)	125 (78%)
CF	62	2 (3%)	43 (69%)
COPD	80	44 (55%)	69 (86%)
DPLD	94	34 (36%)	75 (80%)
LC	80	20 (25%)	38 (48%)
ELD	16	9 (56%)	10 (63%)
PH	51	3 (6%)	51 (100%)
Pneumonia	95	54 (57%)	88 (93%)

4. Discussion

With this study, we were able to define ten core datasets with over 700 essential parameters. During meetings with disease area domain experts, we constantly faced the problem that different participants had different understandings of what "core dataset" means. The participants had different background and were often "lead principal investigators" responsible for studies or registers with narrow research focus. Additionally, previous contact to the scientists who performed this study was often related to technical data integration into the DZL central data warehouse. We encountered three different understandings of the concept "core dataset": (1) with focus on the disease area: list of parameters that are the most important for clinical research in the given domain, (2) with focus on a certain study or register: minimal list of parameters that have to be recorded in each single dataset, and (3) with focus on data integration: exclusive list of parameters that have to be delivered to the central data warehouse.

Besides the definition of which data items are to be considered as core dataset parameters, we did not specify how the domain experts should compose those datasets. We found this a reasonable choice since every disease area has its own peculiarities. In retrospect, it allowed us to adapt to variances. For example, one disease area had to rethink the structure of their diagnostics. Another disease area turned out to require two distinct core datasets for two different diseases, which are both included in the same disease area. For one disease area, a core dataset practically already existed. Additionally, it was a common question to ask the study performers about what parameters other disease areas had already picked for their core dataset. Consequently, one disease area adopted the whole "lung function parameters" section.

We found that the methodology of filtering the most relevant parameters from a given study or register is very efficient. Exemplarily, one disease area named four domain experts who each rated the collected data items of a big study with scores from

zero to four. Items with big gaps (e.g., 4-4-1-4) were discussed in plenary and finally, all items with a sum of 15 or 16 were included in the core dataset.

We defined the DZL Core Dataset by picking all parameters that are included in at least two disease area core datasets. It is arguable whether this is the best definition. Alternatives would be a higher threshold, e.g., at least half of all specific datasets, or a separate survey. The presented dataset gives an impression of all parameters that are of interest for cross-disease area lung research. However, the included parameters may be falsely assumed as to be available in every single study or register within the DZL.

The amount of more than 200 parameters that are considered as essential by domain experts, but were previously missing in the metadata repository, is one aspect confirming the importance of this study. Additionally, the compiled datasets were already used as basis for further data integration of studies/registers into the DZL central data warehouse.

The fact that all disease areas consider the parameter "comorbidities" to be important, but that the recorded characteristics vary considerably, leads to the need for further research. Only when a parameter is recorded in a comparable manner, it can be used to gain meaningful cross-disease research results.

5. Conclusions

Finding the common denominator for the most important clinical parameters in a health research area is challenging, given the diverse perspectives and research interests of dozens of stakeholders. To facilitate this process, we developed a profound definition of what to include in a health research core dataset. We exemplarily defined core datasets within the German Center for Lung Research and share our experiences on what to consider during the composition and harmonization process. Our work resulted in ten core datasets that were fully implemented in our metadata repository CoMetaR. For purpose of interoperability, international classifications were referenced where possible. The composition of core datasets is the next step towards more meaningful health data collections and contributes to future cross-site and cross-disease area research projects.

References

[1] Mark D. Wilkinson. The FAIR Guiding Principles for scientific data management and stewardship.
[2] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med 2018;57(S 01):e50-e56. PMID:30016818
[3] Kulla M, Baacke M, Schöpke T, Walcher F, Ballaschk A, Röhrig R, Ahlbrandt J, Helm M, Lampl L, Bernhard M, Brammen D. Kerndatensatz „Notaufnahme“ der DIVI. Notfall Rettungsmed 2014;17(8):671-681. doi:10.1007/s10049-014-1860-9
[4] Piper I, Citerio G, Chambers I, Contant C, Enblad P, Fiddes H, Howells T, Kiening K, Nilsson P, Yau YH. The BrainIT group: concept and core dataset definition. Acta Neurochir (Wien) 2003;145(8):615-28; discussion 628-9. PMID:14520540
[5] Stöhr MR, Helm G, Majeed RW, Günther A. CoMetaR: A Collaborative Metadata Repository for Biomedical Research Networks. Stud Health Technol Inform 2017;245:1337. PMID:29295418

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230243

Building a Disease Knowledge Graph

Enayat RAJABI[a,b,1] and Somayeh KAFAIE[b,c]
[a] *Cape Breton University, Sydney, NS, Canada*
[b] *Center for Applied Intelligent Systems Research, Halmstad University, Sweden*
[c] *Saint Mary's University, Halifax, NS, Canada*
ORCiD ID: Enayat Rajabi https://orcid.org/0000-0002-9557-0043
Somaye Kafaie https://orcid.org/0000-0002-5685-6487

Abstract. Knowledge graphs have proven themselves as a robust tool in clinical applications to aid patient care and help identify treatments for new diseases. They have impacted many information retrieval systems in healthcare. In this study, we construct a disease knowledge graph using Neo4j (a knowledge graph tool) for a disease database to answer complex questions that are time-consuming and labour-intensive to be answered in the previous system. We demonstrate that new information can be inferred in a knowledge graph based on existing semantic relationships between the medical concepts and the ability to perform reasoning in the knowledge graph.

Keywords. Knowledge Graph, Disease Database, Neo4j.

1. Introduction

Many databases and websites have emerged to allow users to explore medical information on the Web and search for a specific medical disorder or disease. With the complex relationships among different disease concepts, drug ingredients and symptoms in a database, searching for information requires browsing and traversing such relationships to find the tailored answer, which is time-consuming and labour-intensive. Providing a system by defining semantic relationships between different concepts facilitates the search and leads to the discovery of new relations. Hence, many organizations construct a semantic network or knowledge graph to include information about concepts, events and relationships in a graph and perform reasoning [1]. Semantic relationships between entities in a knowledge graph create new concepts and a new level of understanding, allowing machines to make new connections between entities and build new knowledge. Knowledge graphs also have significantly impacted many AI-related applications and information retrieval systems. They have been widely used in healthcare and biomedical systems in applications such as finding relationships between diseases and recommending drugs to patients. They are also utilized in studying genomes and pharmaceutical applications to identify new properties of drugs by predicting drug-drug interactions. This process can help identify a new disease treatment.

Several disease database systems, such as DISNET [2] and DisGeNET [3], leverage semantic relationships between disease concepts, signs, drugs, symptoms and

[1] Corresponding Author: Enayat Rajabi, 1250 Grand Lake Rd, Sydney, NS, Canada; E-mail: enayat_rajabi@cbu.ca.

diagnostic tests associated with a disease to retrieve knowledge from PubMed and Wikipedia. Rather than creating a knowledge graph, extracted data in such systems is usually stored in a database. In this study, we construct a knowledge graph to visualize disease-gene associations, human disease, medications, symptoms and signs based on a cross-referenced disease database. We demonstrate the efficiency of this approach by designing some questions and providing reasoning on the constructed graph. Our results show that knowledge graph reasoning can accelerate identifying critical clinical discoveries and help infer missing facts from existing ones.

2. Method

A knowledge graph can be constructed in different steps, such as data and knowledge acquisition, knowledge enrichment, knowledge storage, and retrieval. The primary data and knowledge acquisition resources include unstructured, semi-structured or structured data. It can be performed based on entity and relation extraction. In the knowledge graph construction process, either a database is constructed or the knowledge graph is created using pre-existing databases. This is the stepping stone towards having information in a machine-readable format, laying the foundation for semantic models, such as ontologies, understanding and using the existing vocabularies, and mapping relationships to add context and meaning to different and distinct data. After mapping, the second step is visualizing the knowledge graph to access a graph-based data view. It allows us to retrieve data, explore the knowledge graph, and write our questions as queries. The following subsections will explain the disease database and the knowledge graph construction process.

2.1. Disease Database

The Diseases Database[2] provides a system for extracting disease-gene associations from biomedical abstracts. It is a cross-referenced index of human disease, medications, symptoms, signs, abnormal investigation findings, etc. This database presents a medical textbook-like index and search portal covering internal medical disorders, symptoms and signs, congenital and inherited disorders, infectious diseases and organisms' medications, common hematology, and biochemistry investigation abnormalities. The database owner provided a subset of 9,400 disease concepts (e.g., diseases, symptoms, and drugs) with 45,507 relationships (e.g., cause-effect, drug family, and risk factor) to us in a Comma Separated Values (CSV) format for data analysis and graph construction.

2.2. Graph Construction

We imported the dataset into a relational database (MySQL) and cleaned it by removing unnecessary columns (e.g., IDs). Table 1 shows a sample of records of the final created dataset before knowledge graph construction. Each disease concept in the dataset is considered a disease node denoted as *n,* related to another concept (*m*) using a relationship *r*. Each dataset record in the database is shown as subject (*n*), predicate(*r*),

[2] http://www.diseasesdatabase.com/ , accessed on 2023-03-12.

and object (*m*) based on the Semantic Web notation[3]. For example, "Phenprocoumon interacts with Paracetamol" is denoted as (Phenprocoumon (*n*), interacts with (*r*), Paracetamol(*m*)) in the knowledge graph. We leveraged the Neo4j Desktop software4 to construct the disease graph and define relationships between the nodes. Neo4j is a graph data platform to store, handle, and query highly connected data in a dataset. Using the LOAD CSV command, we imported the dataset into Neo4j and visualized the knowledge graph. The final disease knowledge graph had 9,400 subject nodes and 3,913 object nodes. The software inferred more than 45,000 relationships.

Table 1. A sample of disease database records

Item-1	Relation	Item-2
Ethanol	may cause	Metabolic acidosis
Amantadine	may cause	Livedo reticularis
Aflatoxins	is a risk factor for	Liver cancer
Respiratory acidosis	is a subtype of	Acidosis

2.3. Query Construction

We followed a question-answering template proposed by [4] for constructing queries with various complexity levels. We also asked a medical doctor to design three questions that require logical reasoning (e.g., hierarchical or inverse) over the subgraphs of the knowledge graph. We later converted the questions to the Neo4j Cypher queries[5] and visualized them in the software. Two types of questions (entity-based or relation-based) were defined. The entity-based question should include at least one relationship between the disease concepts. For example, "What does Dysmenorrhoea cause?" looks for the "cause" relationship between "Dysmenorrhoea" and another entity. To answer this question, an end-user should explore the disease website in different categories to identify the Dysmenorrhoea causes. The relation-based questions use numerous relations and some annotators with various logical operators such as AND, OR, and NOT. The relation between subject and object might be hierarchical or inverse in this type. For example, "Which drug from the '*Folic acid antagonists*' category may cause '*Hepatic failure*' "?

3. Results

The designed questions, the Cypher queries and the results are shown below:

- Question: What diseases are related to "Anemia"? *(Subject: Unknown, Relation: Unknown, Object: "Anemia"?*
  Cypher query: `MATCH (n:First_Item)-[:Relation{type: "may cause"}]-> (b) WHERE b.Name =~"Anemia" RETURN n, b`

[3] https://www.w3.org/standards/semanticweb , accessed on 2023-03-12.

[4] https://neo4j.com/ , accessed on 2023-03-12.

[5] https://neo4j.com/docs/getting-started/current/cypher-intro/ , accessed on 2023-03-12.

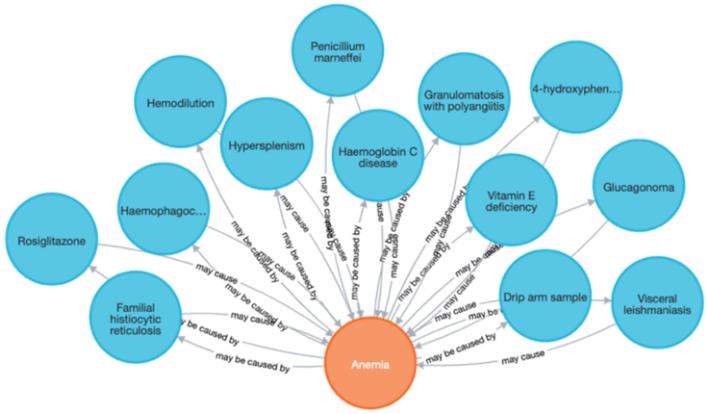


Figure 1. The result of Query 1.

- Question: What Antibiotics drugs may cause "Hepatic failure"? *(Subject: Unknown (a subcategory of Antibiotics drugs), Relation: "may cause", Object: "Hepatic failure")*
  Cypher query:
```
MATCH (n:First_Item)-[c:Relation{type:'may cause'}]->(b:Second_Item{Name:'Hepatic failure'})
MATCH (n)-[dr:Relation{type:'belongs to the drug family of'}]-> (f{Name:'Antibiotics'}) RETURN n, c, b
```

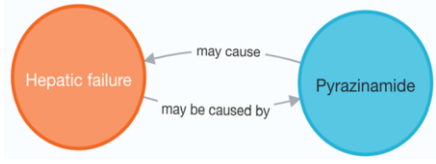


Figure 2. The result of Query 2.

- Question: What are the causes and risk factors of "Acne vulgaris"? *(Subject: Unknown, Relation: ("may cause" and "risk factor for"), Object:"Acnevulgaris")*.
  Cypher query:
```
MATCH (n:First_Item)-[:Relation{type:'may cause'}]-> (b)WHERE b.Name =~ "Acne "+"(?i).*"
MATCH (f)-[dr:Relation{type:'is a risk factor for'}]->(b)RETURN n, f, b
```

Figure 3. The result of Query 3.

The results of Neo4j queries our manual search on the disease database website. To validate the results, we manually explored the disease database website and traversed the relationships between the disease items to retrieve the answers. For example, for the question: What ”Antibiotics” drugs may cause “Hepatic failure”?, we first searched the disease website for “Hepatic failure” keywords. Then, we selected the “may be caused by” relationship and browsed the entities listed there to find their categories and subcategories. If the category was “Drug” and the subcategory was “Antibiotics”, we checked the drug names with the No4j answer.

4. Discussion and Conclusion

Constructing a knowledge graph for a database allows for connecting various concepts semantically. In healthcare, knowledge graph provide a system to explore the connections and discover indirect relationships among diseases, drugs, symptoms and other entities. Our study found some relationships in the disease knowledge graph that were missing from the disease database website. For example, “Acne vulgaris” may be caused by “Propionibacterium acnes” based on the semantic relationships between these two items in the graph; however, the website does not have this information. From a practical point of view, the presented system can be improved by providing a query interface for health professionals to write questions in natural language, as writing Cypher queries might be challenging for non-technical users. In recent years, some studies have started working on the topic [5]; however, it needs a more thorough investigation.

Acknowledgement

This study was possible with the assistance of Dr. Malcolm Duncan, Medical Object Oriented Software Ltd, who shared the Diseases Database with us and answered our technical questions regarding the database. Also, this research has been partially funded by the NSERC Discovery Grant (RGPIN-2020-05869) in Canada and the CAISR Health (Knowledge Foundation grant 20200208 01H) in Sweden.

References

[1] L. Ehrlinger and W. W¨oß. Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS), 48(1-4):2, 2016.

[2] G. Lagunes-Garcia, A. Rodriguez-Gonzalez, L. Prieto-Santamaria, E. P. Garcia del Valle, M. Zanin, and
E. Menasalvas-Ruiz. Disnet: a framework for extracting phenotypic disease information from public sources. PeerJ, 8(e8580), 2020.

[3] J. Pi˜nero, J. M. Ram´ırez-Anguita, J. Sa¨uch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, and L. I. Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Research, 48(D1):D845–D855, 11 2019.

[4] A. Saha, V. Pahuja, M. Khapra, K. Sankaranarayanan, and S. Chandar. Complex sequential question answering: Towards learning to converse over linked question answer pairs with a knowledge graph. In Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.

[5] C. Sun et al. A natural language interface for querying graph databases. PhD thesis, Massachusetts Institute of Technology, 2018.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230244

Performance Benchmarking of FHIR Terminology Operations in ETL Jobs

Joshua WIEDEKOPF[a,b,1], Cora DRENKHAHN[a,b] and Josef INGENERF[a,b]
[a] *IT Center for Clinical Research, University of Lübeck, Germany*
[b] *Institute for Medical Informatics, University of Lübeck, Germany*

Abstract. Interoperability in healthcare cannot be achieved without mapping local data to standardized terminology. In this paper, we investigate the performance of different approaches for implementing HL7 FHIR Terminology Module operations using a benchmarking methodology, to gather evidence on the benefits and pitfalls of these methods in terms of performance from the point-of-view of a terminology client. The approaches perform very differently, while having a local client-side cache for all operations is of supreme importance. The results of our investigation show that careful consideration of the integration environment, potential bottlenecks, and implementation strategies is required.

Keywords. Benchmarking; Controlled Terminologies, Ontologies, and Vocabularies; Interoperability; HL7 FHIR

1. Introduction

The stated goal of current large-scale initiatives in medical informatics, such as the German Medical Informatics Initiative (MI-I) [1] or the European Health Data Space (EHDS) [2] is to make the data held by the large number of stakeholders more accessible to care and research alike. This requires representing medical information in standardized formats, such as the Core Data Set of the MI-I [3], but is inevitably held back by the heterogenous primary systems, so mapping operations to the standardized formats are an essential part of these initiatives. For this, hospitals use Extract-Transform-Load-Pipelines (ETL pipelines) or stream-based processes, which populate the target systems.

Alongside a harmonization of data structures, this requires terminological mappings, whereby coded data elements from the primary system are mapped to the terminology specified in the target dataset. For integration of terminological operations into client code, such as ETL jobs, the HL7 FHIR specification provides the Terminology Module [4]. This specification defines several interactions that should be implemented by HL7 FHIR-based terminology servers (TS) [5–7].

Although there has been an enthusiastic adoption of these servers [7], little guidance about terminological service integration has been published so far, as evidenced by a search on PubMed for the string "(fhir) AND (terminology server)". In particular, performance considerations have not yet been examined in the literature, and in high-

[1] Corresponding Author: Joshua Wiedekopf, Universität zu Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; Telephone: +49 451 3101-5646; E-Mail: j.wiedekopf@uni-luebeck.de.

throughput applications such as ETL pipelines, small benefits in performance of one component can have dramatic impacts on the overall flow rate through the job. This paper, in a semi-quantitative fashion, aims to address this, by comparing several implementation approaches using a benchmarking harness.

2. Methods

This work has been developed in conjunction with an ETL process used in the University Medical Center Schleswig-Holstein (UKSH) for the population of the Smart Infection Control System (SmICS) developed in the HiGHmed consortium [8] of the MI-I. In this ETL process, several local terminologies and coding lists had to be mapped to the standardized representation of the target data structure.

To this end, a rather naïve implementation had been used at first, which downloaded the terminologies from a specified terminology server at start-up, and then executed the needed operations (the *CodeSystem/$lookup* and *ConceptMap/$translate* operations specified by FHIR) via these downloaded files. While the chosen implementation was severely lacking in performance (and later replaced using a faster implementation), we felt that issuing HTTP requests to a terminology server, even in the presence of caching, would present an even greater bottleneck.

As such, we decided on a few alternative strategies that would be evaluated using a benchmarking setup, to try and answer these questions:

- How big of a bottleneck are network operations, and what about IO operations?
- How much of an impact does caching results locally have?
- How "bad" do "naïve" implementations perform compared to "sophisticated" alternatives?

As we decided on using benchmarking for this evaluation, we implemented a separate benchmarking project to our ETL job and used the terminological data of that ETL job as the basis of our benchmarks, instead of using real patient data. Since the ETL job currently only requires the use of the *ConceptMap/$translate* and *CodeSystem/$lookup* operations, only these were benchmarked.

. We used the Java Microbenchmarking Harness (JMH) [9] for generating best-practice [10] benchmarks automatically from our Kotlin code. We used almost entirely default options of JMH to get reliable results: five *warmup iterations* (which don't feed into the overall result), five measurement iterations, each running for 10s, repeating for two (instead of the default five) forks, i.e., ten measurement iterations per implementation. JMH then ran each benchmark as often as possible within the ten-second iteration and reported averages and confidence intervals for the achieved operations per second. Our test data included approximately 1000 concepts for each operation, from which we sampled a random data point for every benchmark call. For implementations, we decided on using these approaches:

- **File**: The original, disk IO-bound file-based approach
- **RAM HashMap**: We read all concepts and mappings into RAM.
- **HTTP**: The operations were delegated to a FHIR-based terminology server, Ontoserver [6], running in the LAN on a dedicated machine (Lenovo ThinkCentre M93p, Intel i5-4570T, 16 GiB RAM with 8 GiB allocated to Ontoserver), connected via Gigabit Ethernet.

- **Redis/SQLite**: The concepts and mappings were read into a local Redis instance (via Docker), a Key-Value-Store, resp. into a local SQLite relational database.

We also benchmarked our input pipeline to make sure that it didn't present a bottleneck itself. Additionally, we implemented several caching approaches:

- **Redis**: We used the same Redis instance as a cache
- **Caffeine**: An open-source in-memory cache for the JVM
- **HashMap/JSON-HashMap**: A cache we implemented ourselves, the first version storing actual Java/Kotlin objects ("POJOs"), the second using Jackson for serialization to and from JSON.
- **Server-side**: For the HTTP requests, we used a Nginx cache in front of the Ontoserver container but didn't use client-side caching.
- **Ktor**: The default client-side cache implementation of the Ktor HTTP framework, with no server-side cache.

All benchmarks were run in a home-office setting, using a laptop computer dedicated to the purpose (Lenovo ThinkPad T490s, Intel i7-8565U, 16 GiB RAM, running Manjaro Linux with no GUI loaded); with computers connected over Gigabit LAN.

3. Results

The results of the evaluation are shown in Figure 1. The graphic features two rows, one for each of the lookup and translate operation, and is split into six horizontal sections for each implementation, the first column showing the input pipeline as the baseline. Every horizontal section shows cache implementations on the X-axis, while the graphs share a logarithmic Y-axis showing operations per second (*ops/s*).

From this visualization, the following points can be inferred:

- Approaches exhibit significant differences in performance, with a range of five orders of magnitude.
- The slowest approach is querying the Ontoserver via HTTP, while loading everything into RAM is the fastest.
- Caching significantly affects performance for all approaches, but the effect of different caches is less pronounced than the effect of the presence of caching itself (c.f. *File*).
- The very naïve file-based approach is indeed more performant than HTTP requests, but not by much. However, caching aids performance significantly, pushing it into the realm of Redis operations.
- SQLite and Redis compare similarly, but Redis seems to suffer from a network latency bottleneck, since the benchmarking code issues network requests to the local Docker container, making performance worse than SQLite.
- Translate operations are a bit more expensive than lookup operations, but the overall pattern of the graphs is nearly identical.

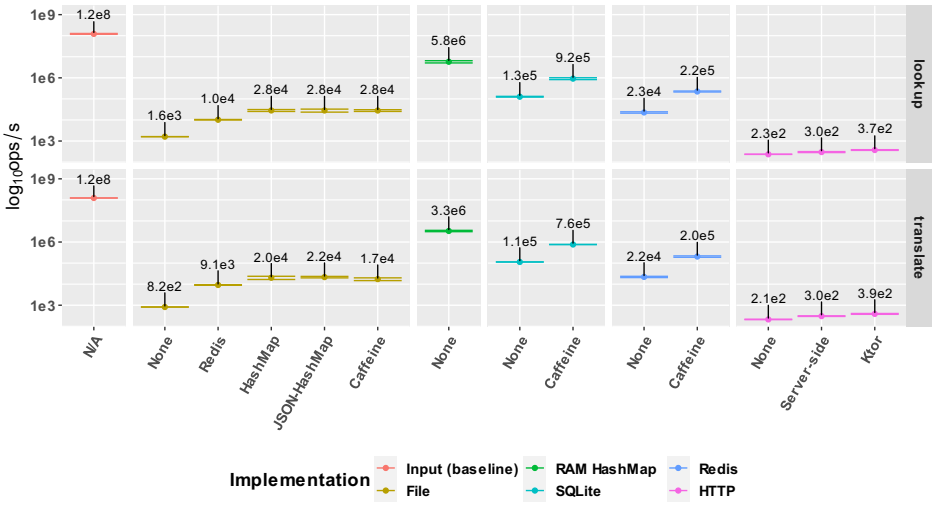


Figure 1. The benchmarking results; for six implementations (tiles) and two operations (rows), each with several cache implementations (X-axes) in logarithmic operations per seconds (Y-axes). The 99.9% confidence interval is shown by the error bars. The figures shown above the error bars are rounded to two significant figures.

4. Discussion

4.1. Implications for the field

The benchmark results show that care must be exercised when integrating FHIR terminology operations in an ETL workflow to avoid performance bottlenecks due to slow IO or network operations. These bottlenecks could then easily translate into much lower throughput of the entire ETL job, due to the importance of terminology operations for interoperability.

Going further, the results illustrate a need for consistent and stable single-source-of-truth for terminology operations, since local operations seem to be more advisable from a performance standpoint than remote operations, as the latter are a severe bottleneck across the LAN and would restrict flow even further across a higher-latency WAN. Hence, implementers must be able to access an up-to-date feed of the relevant terminology resources available in their jurisdictions that can then be loaded in local processes. Steps toward these kinds of infrastructures are being taken in many jurisdictions, greatly improving on the status-quo.

4.2. Limitations

The biggest limitation of this study is arguably the evaluation of only two operations, *CodeSystem/$lookup* and *ConceptMap/$translate*. We deliberately restricted us to these operations, since implementing them from a given set of FHIR resources is quite simple, limiting the possibility of severe implementation errors, which would render the benchmark meaningless. Operations like subsumption testing and the handling of more complex terminologies is far from trivial and requires extreme care.

In this way, the results of the study should not be misunderstood as a statement against the use of terminology servers, which have shown the worst performance in our tests. The operations we have examined are a realistic simulation of a limited-scope ETL job, but that will not be the case for all types of ETL jobs. For more complex operations, terminology servers remain the only option.

The quantitative results of this study should not be used to make major architectural decisions on their own. Instead, ETL developers and architects should consider the requirements and circumstances of their environment carefully. For more complex operations, it is likely that a hybrid setup, whereby complex operations are carried out by a local terminology server, and simpler operations by a local database, performs best. To achieve the best performance possible, performing benchmarks on the production infrastructure are advisable.

5. Conclusions

We analyzed several approaches for implementing terminological operations, varying in implementation and conceptual complexity, using a benchmarking approach. We maintained the point-of-view of ETL integration in our experiments since the performance of the different approaches is especially relevant in this regard. The results show a marked difference in the operations per second achievable by the different approaches, highlighting the importance of caching, and the relevance of careful consideration and architectural decisions.

Acknowledgments: This work is funded by the German Federal Ministry of Education and Research (BMBF) as part of the Medical Informatics Initiative Germany, grant 01ZZ2312A (SU-TermServ).

References

[1] Semler SC, et al. German Medical Informatics Initiative. Methods Inf Med 2018;57:e50–6.
[2] European Commission. A European Health Data Space: harnessing the power of health data for people, patients and innovation 2022. https://health.ec.europa.eu/system/files/2022-05/com_2022-196_en.pdf.
[3] Ganslandt T, Boeker M, Löbe M, Prasser F, Schepers J, Semler S, et al. Der Kerndatensatz der Medizininformatik-Initiative: Ein Schritt zur Sekundärnutzung von Versorgungsdaten auf nationaler Ebene. Forum Med-Dok Med-Inform 2018;20:17–21.
[4] HL7 International. Terminology Module – HL7 FHIR R4 Specification 2019. https://www.hl7.org/fhir/terminology-module.html (accessed March 24, 2021).
[5] Benson T, Grieve G, editors. Principles of Health Interoperability: FHIR, HL7 and SNOMED CT. Fourth edition. Cham: Springer International Publishing; 2021.
[6] Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. J Biomed Semant 2018;9:24.
[7] Kleinoscheg G, Tanjga N, Svec N, Rainer-Sablatnig S, Sabutsch S. TerminoloGit - An Open Source Terminology Server for Large Scale eHealth Environments. Stud Health Technol Inform 2022;293:79–84.
[8] Wulff A, Baier C, Ballout S, Tute E, Sommer KK, Kaase M, et al. Transformation of microbiology data into a standardised data representation using OpenEHR. Sci Rep 2021;11:10556.
[9] OpenJDK. Java Microbenchmarking Harness (JMH). GitHub 2022. https://github.com/openjdk/jmh (accessed January 2, 2023).
[10] Horký V, Libič P, Steinhauser A, Tůma P. DOs and DON'Ts of Conducting Performance Measurements in Java. Proc. 6th ACMSPEC Int. Conf. Perform. Eng., New York, NY, USA: Association for Computing Machinery; 2015, p. 337–40.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230245

Fitness for Use of Anatomical Therapeutic Chemical Classification for Real World Data Research

Ines REINECKE[a,1], Elisa HENKE[a], Yuan PENG[a], Martin SEDLMAYR[a] and Franziska BATHELT[a]
[a] *Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany*

Abstract. *Introduction:* Real-world data (RWD) is gaining importance in research. For instance, the European Medicines Agency (EMA) is currently in the process of establishing a cross-national research network that utilizes RWD for research. However, data harmonization across countries must be carefully considered to avoid misclassification and bias. *Objectives:* This paper aims to investigate the extent to which a correct assignment of RxNorm ingredients is possible for medication orders that include only ATC codes. *Methods:* In this study, we analyzed 1,506,059 medication orders from the University Hospital Dresden (UKD) and merged them with the ATC vocabulary in the Observational Medical Outcomes Partnership (OMOP) including relevant relationship mappings to RxNorm. *Results:* We identified 70.25% of all medication orders were single ingredients with direct mapping to RxNorm. However, we also identified a significant complexity in mappings for the other medication orders that was visualized in an interactive scatterplot. *Discussion:* The majority of medication orders under observation (70.25%) are single ingredients and can be standardized to RxNorm, combination drugs pose a challenge due to the different approaches of ingredient assignments in ATC and RxNorm. The provided visualization can help research teams gain a better understanding of problematic data and further investigate identified issues.

Keywords. OHDSI, OMOP, RxNorm, ATC, interoperability, medication, RWD

1. Introduction

Real-world data (RWD) is health data that is routinely collected from multiple data sources during treatment and offers new research opportunities complementary to randomized controlled trials. Digitization is leading to the availability of electronic health records and are the catalyst for establishing retrospective research with RWD. Huge efforts are currently undertaken in Europe to benefit from RWD in research by establishing large, multi-national networks such as demonstrated by the European Health Data Evidence Network (EHDEN) [1]. Additionally, the European Medicines Agency (EMA) is currently in the process of establishing the Data Analysis and Real World Interrogation Network (DARWIN) to access and analyze healthcare data from across the

[1] Corresponding Author: Ines Reinecke, Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany; E-mail: ines.reinecke@tu-dresden.de.

European Union with the goal of running large RWD studies to support regulatory decision-making [2]. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), developed by the Observational Health Data Sciences and Informatics (OHDSI) community, is used by EHDEN and DARWIN and is becoming increasingly important for retrospective research with RWD [3]. A major challenge in harmonizing RWD from different sources and countries into a common data model is the transition of non-standardized data and terminologies into a standardized format and terminologies. However, these activities are necessary to ensure semantic interoperability within cross-country research networks [4,5]. The translation activity is subject of bias due to the risk of misclassification of clinical facts such as drug exposures because of multiple mapping options [6]. Medication data are very heterogeneous and need to be harmonized for the use in the OMOP CDM. It is necessary to use the mandatory standard terminology RxNorm. RxNorm is a standardized nomenclature for all clinical drugs in the United States provided by the U.S. National Library of Medicine [7]. Unfortunately, hospital information systems often document drug exposure as free text or with ingredient information only as Anatomical Therapeutic Chemical Classification (ATC) codes [8]. A major difference between the active ingredients in ATC and RxNorm is that a single ATC code can combine more than one active ingredient in one code, whereas in RxNorm each active ingredient is represented by a separate code [9]. Therefore, this paper reports on the evaluation of the fitness for use of ATC for research in the OMOP CDM by checking to what extent the correct active ingredient in RxNorm can be assigned unambiguously. We present a visualization solution that helps interdisciplinary teams to gain better understanding of their medication data in OMOP CDM, when source data only contains ATC codes and mapping to a standard terminology is required.

2. Methods

This work is based on the medication order data from the University Hospital Carl Gustav Carus Dresden (UKD) from 2016 to 2020. It consists of 1,768,153 medication orders that have been already analyzed and improved in a previous work as described by Reinecke et al. [8]. For the present work, a subset of 85.18% (1,506,059) medication orders that has an ATC Level 5 code assigned, is used as input data. The most recent ATC vocabulary version from September, 07 2021 (version flag RxNorm 20210907) from the OHDSI ATHENA vocabulary service [10] containing 6,497 valid concepts has been imported into an OMOP database version 5.3.1. The required ATC to RxNorm vocabulary relationship information has been exported from the *concept* and *concept_relationship* tables in OMOP CDM for further exploration utilizing SQL statements. This export has been limited to the relationships of interest ("ATC -RxNorm pr up", "ATC -RxNorm pr lat", "ATC -RxNorm sec lat", "ATC -RxNorm sec up") since those represent the relationships for active ingredients available as ATC codes and their related RxNorm ingredients in accordance to the official OHDSI ATC vocabulary documentation [11]. In a first step, the relationship information has been enriched to provide a single row for each ATC Level 5 code with at least one of the above-mentioned relationship types containing the number of relationships types each calculated as numerical value in a separate column. Additionally, the total number of existing relationships between ATC Level 5 code and RxNorm has been calculated and added into an additional column. Second, the medication orders have been transformed according to the *drug_exposure* table in the OMOP CDM. Third, medication data in the

OMOP format has been merged with the transformed relationship information outcome from the first step. The three above described steps allow the determination of all ATC codes with a relationship to one or more RxNorm codes and show the challenges due to multi mappings during the transformation of medication orders. Since numerical results are rather confusing due to the high number of medical orders containing combination drugs, we developed an interactive visualization in addition to the table that contains the results. The visualization is designed interactively with search and hover capabilities to support identification of medication orders containing ATC codes that need special attention due to relationship complexity and high occurrence. Data analyzes and visualization has been implemented in Python Version 3.9.1 utilizing the following libraries: Pandas, Matplotlib and Bokeh. The complete source code and the interactive visualization can be accessed and downloaded on Zenodo [12].

3. Results

We identified 4811 ATC codes of the vocabulary (71.38%, 4811/6479) that have at least one mapping relationship to any of the 4 investigated types. The number of relationship mappings per ATC code can be large as exemplified in Table 1 for the 5 most frequent mapping relationship type combinations from a total of 363 different mapping combinations. There are 3655 (56.41%, 3655/6479) ATC codes with a single ingredient mapping to a RxNorm ingredient. ATC codes that contain combination ingredients can still have an exact mapping of ingredients to RxNorm, e.g.: J05AR13= "lamivudine, abacavir and dolutegravir; systemic" with relationships to the exact 3 RxNorm ingredients "abacavir", "dolutegravir" and "lamivudine". But there are more generic ATC codes with the primary ingredient mapped to RxNorm and additional ingredients unknown, e.g. ATC code C09BA05= "ramipril and diuretics" with a mapping to the RxNorm ingredient "ramipril" and 30 options for the diuretics. Thus, a specific mapping without the drug product information is not possible.

Table 1. Most frequent mapping combinations for ATC to RxNorm ingredient relationships

Number of ATC codes (exist in RWD)	ATC RxNorm pr lat	ATC RxNorm pr up	ATC RxNorm sec lat	ATC RxNorm sec up	# of Medication orders (%)
3655 (567)	1	0	0	0	1,057,986 (70.25)
143 (10)	1	0	1	2	8,859 (0.59)
95 (17)	2	0	0	0	28,717 (1.91)
55 (0)	1	0	2	3	0 (0.00)
34 (4)	1	0	1	3	2,717 (0.18)

The merge of the medication orders with the existing mapping relationships lead to a total of 567 ATC codes in our data with exactly one relationship to RxNorm. This identifies 70.25% (1,057,986/1,506,059) of all medication orders as single ingredients that can be mapped directly to a single RxNorm ingredient. A visualization with a search option, a result table and a scatterplot was developed (Figure 1). The search field offers filtering capabilities for ingredient names or ATC codes. The result table shows details on the ATC codes including its frequency in the medication order data, total number of mapping relationships to RxNorm and the ingredient name. The scatterplot represents each ATC code occurred in the medication orders as a single dot. The frequency of each ATC code in the medication order data is shown on the x-axis and the level of mapping

complexity to RxNorm is represented on the y-axis. For example, ATC code B05BB01 was prescribed 74,314 (frequency) times and has a total of 349 different mapping relationship types to RxNorm. This interactive visualization is used together with the pharmacist experts at the UKD to investigate the data further.

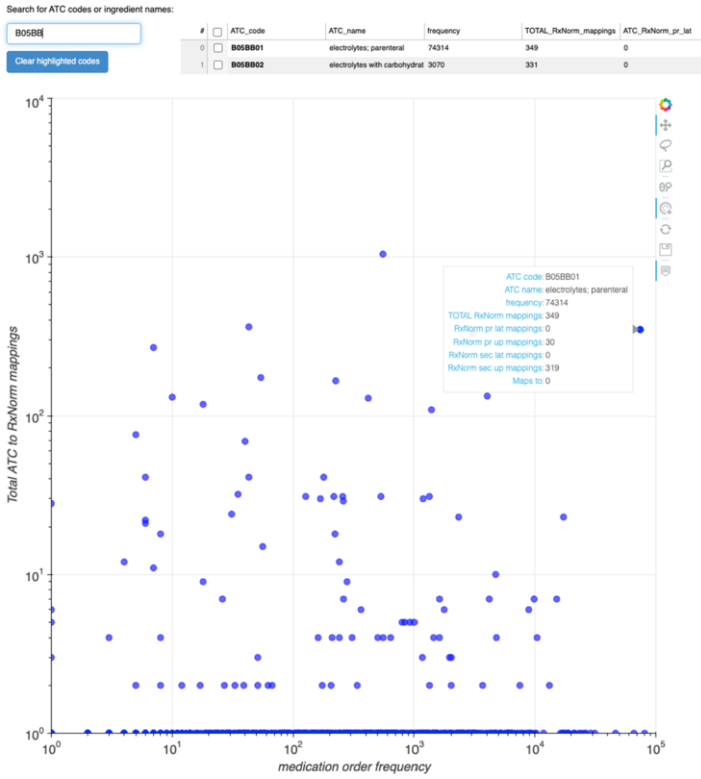


Figure 1. Medication orders by frequency and number of mapping relationships

4. Discussion and Conclusion

In this paper we were able to apply RxNorm ingredients to the majority of the medication orders at the UKD and thus enable research on OMOP CDM for single ingredient drugs. However, ATC is challenging when a single ATC code represents ingredient combinations. For some ATC codes a large number of mapping relationships to the RxNorm were identified (Figure 1) that cannot be assigned without further knowledge about the medication orders to the appropriate standard terminology required by the OMOP CDM. The interactive visualization created in this work was applied to only one site. The developed visualization can be used by other research teams, because our implementation uses the OMOP CDM as an input format. A prerequisite is the presence of the vocabulary in ATHENA and the mappings to a standardized terminology. In a next step we are going to analyze study protocols of studies executed by the OHDSI community teams to check how often ATC codes with combination ingredients have been used for cohort definitions and to answer research questions in the past years.

Declaration

Conflict of Interest: The authors declare that there is no conflict of interest.
Funding: This work is part of the MIRACUM project funded by the German Ministry of Education and Research (FKZ 01ZZ1801A/L).

References

[1] EHDEN, EHDEN - European Health Data & Evidence Network, (n.d.). https://www.ehden.eu/ (accessed January 2, 2023).
[2] European Medicines Agency (EMA), Data Analysis and Real World Interrogation Network (DARWIN EU), (n.d.). https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu (accessed January 2, 2023).
[3] Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The Usage of OHDSI OMOP–A Scoping Review. German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal. 2021:95-103. doi:10.3233/SHTI210546.
[4] Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. International Journal of Medical Informatics. 2023 Jan 1;169:104925. doi:10.1016/j.ijmedinf.2022.104925.
[5] Puttmann D, De Keizer N, Cornet R, Van Der Zwan E, Bakhshi-Raiez F. FAIRifying a Quality Registry Using OMOP CDM: Challenges and Solutions. InChallenges of Trustable AI and Added-Value on Health 2022 (pp. 367-371). IOS Press. doi:10.3233/SHTI220476.
[6] EMA, A common data model in Europe? – Why? Which? How?, (2018). https://www.ema.europa.eu/en/documents/report/common-data-model-europe-why-which-how-workshop-report_en.pdf (accessed January 20, 2023).
[7] National Institutes of Health (NIH), National Library of Medicine - RxNorm, (n.d.). https://www.nlm.nih.gov/research/umls/rxnorm/index.html (accessed January 2, 2023).
[8] Reinecke I, Siebel J, Fuhrmann S, Fischer A, Sedlmayr M, Weidner J, Bathelt F. Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation. JMIR Medical Informatics. 2023 Jan 25;11(1):e40312. doi:10.2196/40312.
[9] Bodenreider O, Rodriguez LM. Analyzing US prescription lists with RxNorm and the ATC/DDD Index. InAMIA Annual Symposium Proceedings 2014 (Vol. 2014, p. 297). American Medical Informatics Association.
[10] ATHENA standardized vocabularies – OHDSI, (n.d.). https://www.ohdsi.org/analytic-tools/athena-standardized-vocabularies/ (accessed January 17, 2022).
[11] OHDSI, Vocabulary ATC description, (n.d.). https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:atc.
[12] Reinecke I. Source code of the visualization, (2023). https://zenodo.org/record/7521226 (accessed January 10, 2023).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230246

Observational Cohort Study Dedicated to Autism Spectrum Disorder: Milestone Steps, Results Updates, Perspectives

Mohamed BEN SAID[a,1], Laurence ROBEL[b], Fanny KPENOU[a], Jean-Philippe JAIS[a], Mario SPERANZA[c] and TEDIS Study Group
[a] AP-HP Necker-Enfants Malades Hospital, *Biostatistics Unit, Paris Cité University, Paris, France*
[b] *Dept of Child and Adolescent Psychiatry, GHU Psychiatry and Neuroscience, Paris, France, INSERM UMR 1018 CESP Paris-Saclay University, France.*
[c] *Dept of Child and Adolescent Psychiatry, Versailles Hospital, Versailles, France, INSERM UMR 1018 CESP Paris-Saclay University, France.*

Abstract. Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by persistent difficulties in two domains: social communication and interaction, alongside with restricted, repetitive pattern of behaviors. It affects children and persists into adolescence and adulthood. Its causes and underlying psychopathological mechanisms are unknown and remain to be discovered. TEDIS cohort study developed over the decade 2010-2022, in Ile-de-France region, includes 1300 patients' files up to date, with valuable health information drawn from ASD evaluation. It provides researchers and decision makers with reliable data source to improve knowledge and practice in the context of ASD patients.

Keywords. Austim Spectrum Disorder, TEDIS, Cohort Study, Neurodevelopmental disorder, ASD

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition with an estimated median prevalence of 100/10,000 and a median percentage of autism cases co-occurring with intellectual disability of 32% [1]. ASD are characterized by persistent difficulties in two domains: social communication and interaction, alongside with restricted, repetitive pattern of behaviors [2]. Diagnosis is based upon precise behavioral and communication analysis of children of three-to-five years old [3]. Children and adults with autism can have a happy and healthy life but urgent action is required to promote these outcomes [4]. Expert evaluations of ASD' patients in specialized medical centers in Ile-de-France region called « Centre de Diagnostic et d'Evaluation de l'Autisme (CDE) » and « Plateforme Autisme de Proximité (PDAP) » are organized around in-depth psychopathological assessment besides multidisciplinary domains investigations. It

[1] Corresponding Author: Mohamed Ben Said, Biostatistics Unit, Necker University Hospital. 149 rue de Sèvres. 75015, Paris. France. E-mails : mohamed.bensaid@aphp.fr, mohamed.ben-said@ext.paris-descartes.fr.

generates valuable health information for apprehending underlying psychopathological, physiological processes of the ASD' affection. An information system called TEDIS, was developed to manage such information, in the context of a) ease-of-use in daily practice in clinical settings, b) respect of privacy, ethics, data integrity and security, c) being adaptable to knowledge domain evolutions, e) expandable to new participating centers. This article aims at presenting major milestone steps during the last decade and the solutions brought at each step. Excerpts of data analyses results will be discussed in relation with governance, and collaboration perspectives.

2. Methods

Four major phases punctuate TEDIS evolution during the last decade (2011-2022).

2.1. TEDIS Milestone steps

Phase 1 [2009-2010]: Dedicated to the development of conceptual data model and the application system. TEDIS was implemented in relational MySQL™ database, within an n-tiers architecture web information system, using light-weight client engineering and free-license software: HTML, CSS, JavaScript, Apache-tomcat and Java ™ language.
Phase 2 [2011–2017]: TEDIS system was tested in a stand-alone mode by psychiatrists, psychologists, and medical residents at Necker university hospital.
Phase 3 [2017 – 2019]: Institutional endorsement and support from the regional public health agency in parallel with multi-centers deployment through the Assistance Publique des Hôpitaux de Paris (APHP) Intranet.
Phase 4 [2019 – Now] Strengthen applications' protection of patient private information before Internet deployment in mid-2019. Hosting new participating centers was facilitated but requiring data model evolutions to fit with local clinical practice [5].

2.2. TEDIS data

TEDIS ASD' patient inclusion is based on ICD-10' F84 class codes reported in the patient discharge summary. The main diagnosis points to the reason which mobilized the most of the medical, clinical and care resources. F84 class codes, may also be present in one of three associated diagnoses, referenced usually to enrich patient' medical characterization. A selection of patients of interest subsets, whose diagnoses are uncertain at the moment of the assessment and whose discharge summary include results of two out of three gold standard ASD' assessment tests [2], are included in TEDIS' cohort. The data production workload, followed the chronology of expert centers joining the study. Data quality was ensured by medical residents in phase 2 and in phase 3, then by the clinical research assistant. Main statistical descriptive analyses results are introduced in this paper. They are developed in R statistics software and integrated in an R-Markdown programs, adapted to the cohort evolution and growth.

[2] CARS: Childhood Autism Rating Scale; ADI-R: Autism Diagnostic Interview – Revised; ADOS: Autism Diagnostic Observation Schedule

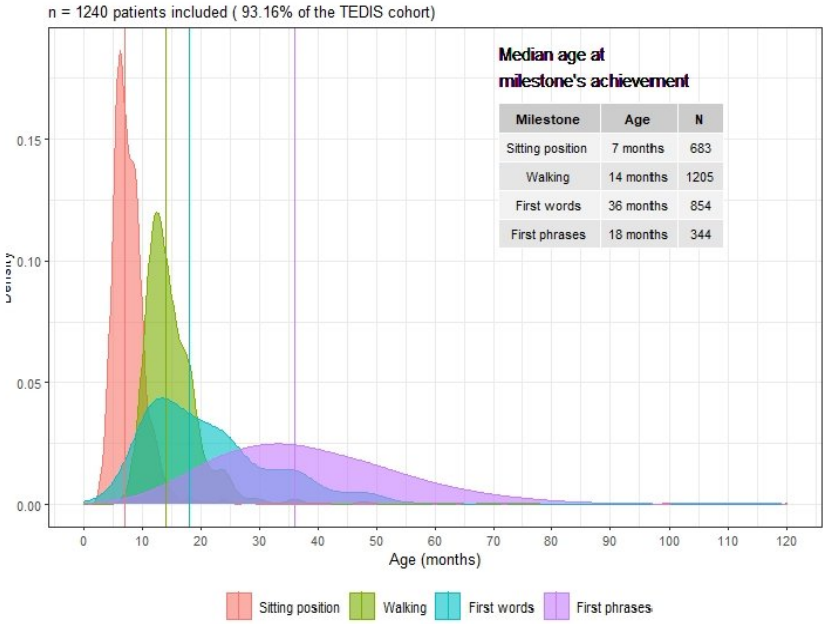


Figure 1. Motor and speech development ages distributions. Documented cases n=1240 with at least one out four information documented: Median age for sitting: 7 months old. Walking median age: 14 months. Median age at first words: 18 months, and median age at first phrases: 36 months old.

3. Results

TEDIS' cohort included as of January 20th 2023, 1331 ASD' patient records. Patients diagnosed with an F84 ICD-10 class code represent 90% of the participants. Patient ratio confirms the four males to one female and the median age of 5 years old for expert' ASD diagnosis. 1255 patients, reside in Ile-de France region. Most of the ASD patients belong to bi-parental family structures: 973 cases (73%). The mean father age is of 36 years old in 944 cases and the mean mother age is of 32 years old in 958 cases. Parents belong to all socio-professional categories, according to the national institute of statics INSEE, with a noticeable proportion of women without professional activity: 30% in both the bi-parental families and the 279 mono-parental families. There are no noticeable psychological difficulties during pregnancy period (18.7% cases) with 157 reported cases among women (including 73 cases of anxiety) and 65 cases including 8 anxiety situations among the fathers. Pregnancies were carried to term in 1024 cases (77%), and the birth parameters are in normal boundaries in the majority of cases (85%). The absence of language capacity is reported in 477 cases, mostly (92%) with F84 diagnosis. Motor and speech median acquisition age results are reported in Figure 1. First wording capacity acquisition age distribution in figure 1, shows the influence of the extreme values (120 months). Feeding and sleeping habits are reported in 726 ASD patient cases of ASD patients, with absence of disorder in 18% cases of sleeping and in 23% cases of feeding habits. Feeding selectivity is observed in 49% of cases and nocturnal awakening are observed in 29% of cases. Achieving toilet control mean age is about 39.1 months. Most of study ASD' patient population, 1163 cases, beneficiated from schooling, ambulatory specific health care and social measure, before ASD expert assessment.

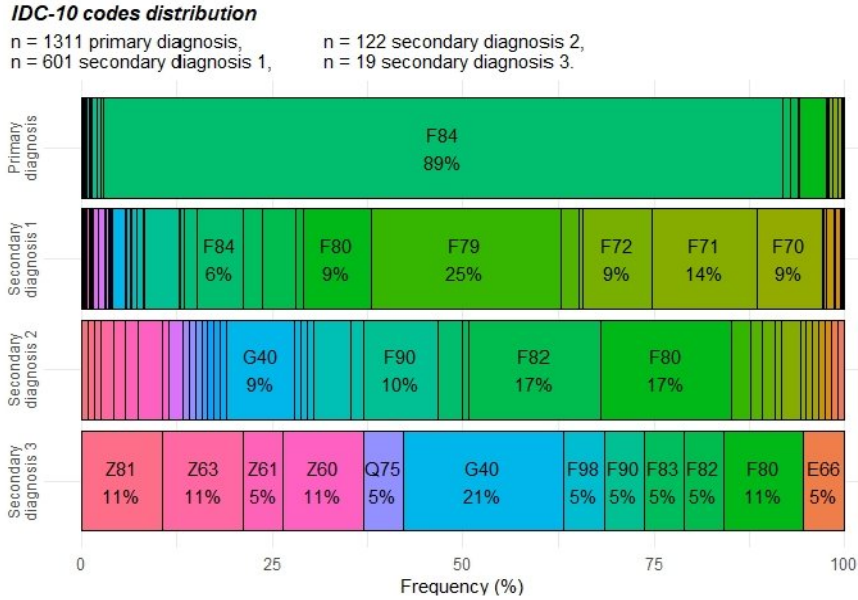


Figure 2. ICD-10 codes distribution in TEDIS' cohort. Documented cases n=1311.89% belong to ASD class coding class F84 for the main diagnosis. In the first associated diagnosis position, F7 class codes of intellectual deficiency represents 57%. In the second diagnosis position there is a mixture of attention deficit-hyperactivity disorders, speech and language disorders, epilepsy and seizures, while in forth associated diagnosis position, we find a mixture of neurological and developmental disorders.

Ambulatory care consisted mainly in speech (21%) and motor (21%) supporting measures and 56% beneficiated from social measures and handicap status recognition opening facilities for free public health care and social accompaniment.

Psychiatry and psychology assessment tests results are reported in 1233 (93%) cases. Reported cases are of 874 for ADI-R, 593 for CARS, 534 for ADOS and ADOS-2, wheschler scales : 221 reported cases. There are a mean of 2.7 Psychiatry and psychology assessment tests for each patient record. Figure 2, shows the main diagnosis dominance of the ICD-10 class codes F84, characterizing ASD' patients. Intellectual deficiency, attention deficiency-hyperactivity and neurological symptomatology are frequently present in ASD' symptomatology. TEDIS' cohort patient data specific subsets, already permitted six medical academic research works and publications in neuro-psychiatry. A similar number of methodological indexed articles were published in medical informatics communities. The presentations to the psychiatrics participating in the study cohort of the of data analysis automatic report were encouraging and motivating for pursuing feeding the cohort with the expert information of ASD patients' assessments.

4. Discussion

TEDIS cohort tries to keep up with daily practice in clinical settings. This objective is challenged by recurrent evolutions in ASD' medical codification and assessment criteria during the last decade, besides new versions of assessment scales and psychometrics testing and variations in assessment practice in the participating centers need. The partnership between clinicians and methodologists and the data quality control by the

clinical research assistant, try to guarantee development, evolution, and maintenance of a reliable data source. Building a structured data model is a long process. It offers however facilities for statistical quantitative and qualitative analyses, allowing rapid advances in producing research results. Participating psychiatrists adopted a governance schema with a steering committee and a scientific committee. A charter commits participants to the cohort study project and aggregated, de-identified patient data sets will be made available for researchers on demand, to be addressed to the scientific committee.

5. Conclusion

Building a reliable clinical data source for research and public health policies remains a challenge, involving domains experts and methodologists. Structured data modeling is helpful for rapidly advancing research and taking advantages from available statistical tools and capacities. Data quality control is evident in modern data source development. Governance and data sharing need to guarantee a safe use of the data with respect of ethics, privacy, scientific objectives, and benefits to the community.

Acknowledgement

Psychiatrists in the centers participating in TEDIS cohort study (TEDIS study group), are fully thanked for their support. The Centre Ressource Autisme Ile de France (CRAIF) and the Agence Regionale de Santé Ile de France (ARS-IDF) are thanked for supporting and funding TEDIS study ASD patient cohort.

References

[1] Zeidan J, Fombonne E, Scorah J, Ibrahim A, Durkin MS, Saxena S, Yusuf A, Shih A, Elsabbagh M. Global prevalence of autism: A systematic review update. Autism Res. 2022 May;15(5):778-790. doi: 10.1002/aur.2696. Epub 2022 Mar 3. PMID: 35238171; PMCID: PMC9310578.

[2] Hyman SL, Levy SE, Myers SM; COUNCIL ON CHILDREN WITH DISABILITIES, SECTION ON DEVELOPMENTAL AND BEHAVIORAL PEDIATRICS. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. Pediatrics. 2020 Jan;145(1):e20193447. doi:10.1542/peds.2019-3447. Epub 2019 Dec 16. PMID: 31843864.

[3] Ashwood KL, Buitelaar J, Murphy D, Spooren W, Charman T. European clinical network: autism spectrum disorder assessments and patient characterisation. Eur Child Adolesc Psychiatry. 2015 Aug;24(8):985-95. doi: 10.1007/s00787-014-0648-2. Epub 2014 Dec 4. PMID: 25471824.

[4] Lord C, Charman T, Havdahl A, Carbone P, Anagnostou E, Boyd B, Carr T, de Vries PJ, Dissanayake C, Divan G, Freitag CM, Gotelli MM, Kasari C, Knapp M, Mundy P, Plank A, Scahill L, Servili C, Shattuck P, Simonoff E, Singer AT, Slonims V, Wang PP, Ysrraelit MC, Jellett R, Pickles A, Cusack J, Howlin P, Szatmari P, Holbrook A, Toolan C, McCauley JB. The Lancet Commission on the future of care and clinical research in autism. Lancet. 2022 Jan 15;399(10321):271-334. doi: 10.1016/S0140-6736(21)01541-5. Epub 2021 Dec 6. Erratum in: Lancet. 2022 Dec 3;400(10367):1926. PMID: 34883054.

[5] Ben Said M, Robel L, Chaste P, Perisse D, Oreve MJ, Zylberberg P, Philippe A, Jousselme C, Lacoste S, Speranza M, Hafsa I, Cisse F, Cherqaoui Z, Jais JP. TEDIS, a Comprehensive Data Model for In-Depth Clinical Assessment of Patients Affected with Neuro-Developmental Disorders Including Autism. Stud Health Technol Inform. 2020 Jun 16;270:1401-1402. doi: 10.3233/SHTI200462. PMID: 32570679.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230247

Concept for a Basic ISO 14721 Archive Information Package for Clinical Studies

Matthias LÖBE[a,1], Alessandra KUNTZ[b], Christian HENKE[b], Frank MEINKE[a], Ulrich SAX[b,c] and Alfred WINTER[a]
[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany*
[b] *Department of Medical Informatics at the University Medical Center Göttingen, Germany*
[c] *Campus-Institute of Data Science (CIDAS), Göttingen, Germany*
ORCiD ID: Matthias Löbe 0000-0002-2344-0426, Alessandra Kuntz 0000-0002-8259-2577, Christian Henke 0000-0002-4541-4018, Frank Meineke 0000-0002-9256-7543, Ulrich Sax 0000-0002-8188-3495

Abstract. Secondary use of medical data for research is desirable for intrinsic, ethical and financial reasons. In this context, the question becomes relevant as to how such datasets are to be made accessible to a larger target group in the long term. Typically, datasets are not extracted ad hoc from the primary systems, because they are processed qualitatively (FAIR data). Special data repositories are currently being built for this purpose. This paper examines the requirements for the reuse of clinical trial data in a data repository utilizing the Open Archiving Information System (OAIS) reference model. In particular, a concept for an Archive Information Package (AIP) is developed with the central focus on a cost-effective trade-off between the effort of creation for the data producer and the comprehensibility of the data for the data consumer.

Keywords. Data sharing, Research Data Management, OAIS, ISO 14721, Clinical trials, Archiving, Data repositories, Data reuse, FAIR

1. Introduction

Data sharing, the reuse of data once collected for further research purposes, has moved from an altruistic attitude of individual researchers to an obligation of many funders or scientific publishers in recent years. Data sharing makes sense for scientific, ethical, and resource reasons. Nevertheless, it is not widely practiced because, first, it involves a significant effort for the data provider that is currently not compensated, second, there is a lack of data repositories for permanent storage and legally secure access (particularly for sensitive personal health data), and third, there are no widely accepted ideas about which data structures are essential for data recipients in complex research projects.

In the field of medical research, local research data centers are currently being set up in Germany that will have the technical (hardware, storage space, network

[1] Corresponding Author: Matthias Löbe, Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany; E-mail: matthias.loebe@imise.uni-leipzig.de.

connection), content-related (data body, metadata), legal (framework agreement, usage regulations, consent, pseudonymization) and organizational (data access board, trust office, ethics committee) prerequisites. In projects such as the Medical Informatics Initiative (MII) [1] or the National Research Data Infrastructure for Personal Health Data (NFDI4Health) [2], staff positions have been created for data management. Thus, (semi-)permanent structures will exist for the storage and retrieval of health data.

On the other hand, these data repositories differ from institutionally funded libraries with a statutory mandate. They have less financial resources and are not primarily intended to archive health data in their entirety for all time, but rather to provide target group-oriented data packages that can be reused for secondary research projects. A number of special requirements arise for the new research data centers:

1. Managing data in a repository creates a third party of data stewards between data producers and data consumers. Data stewards have generally not been involved in the data collection process and cannot answer queries from potential consumers themselves. Therefore, standards must be created for data dissemination packages that make the data they contain *discoverable* and *reusable* by target audiences.
2. In many cases, there will be a time gap between when the data is delivered to the repository and when consumers first request it. It is possible that the data-producing projects have already been completed and no permanent contacts are available anymore. Therefore, standards need to be created for data dissemination packages that make the data *interpretable* and include descriptive, semantic, and provenance metadata.
3. Few current research projects have a financial line item for longer-term data sharing. The agreed standards must therefore also be guided by what is feasible for data producers when handing over the data and for data managers in routine operations, and what risks exist in terms of data protection and data security. Therefore, standards need to be created for cost-effective data dissemination packages that can be assembled by artifacts already existing in routine operation, automatic conversation and limited manual enrichment.

2. Methods

The Open Archival Information System (OAIS) [3] is a reference model for a digital archive, which has also been published as ISO standard 14721 (2012). It describes the components of an archival system for long-term preservation. From a data perspective, OAIS is based on *information packages* that contain descriptive information in addition to the actual content to be preserved. The most important type of information package is the *Archive Information Package* (AIP) (see Figure 1), which represents the underlying archived artifact. In addition to the actual data object, it contains detailed metadata on the structure and interpretation of the content (*Representation Information*) as well as on identifications, context, provenance and access options (*Preservation Description Information*). For the findability of the AIP, additional descriptive information is provided [4].

An expert-based approach was chosen to create a concept for an AIP specific to clinical trials. Unlike traditional archives or libraries, medical research data centers are

usually sub-organizational units of IT departments that also perform other tasks such as data extraction from primary systems or operating Clinical Data Management Systems for data collection in clinical trials. For this reason, they are familiar with both the data-generating processes as well as the needs of the target group and can define the scope of an AIP in collaboration with data managers and investigators. Especially in the environment of regulated research, formal standard operating procedures (SOP) and working instructions exist for all substeps of a clinical trial. The documents and data files described there have been evaluated to determine a) the extent to which they are necessary or useful for understanding the clinical data and the research project, b) whether their content contains sensitive information in terms of data protection or intellectual property, and c) what effort would be involved in generating or transforming them to suit the target group. The last point aims at the fact that documents and data for controlled internal use have to meet lower requirements than externally published assets. Enrichment is often described under the keyword FAIRification Workflow [5] and can be a complex procedure. As previously mentioned, the focus of the work was on generating immediately understandable and cost-effective packages. High demands and mandatory standards of the regulatory authorities (CDISC, FDA, EMA) exist for the execution and analysis of drug approval studies. These were not the aim of the work.

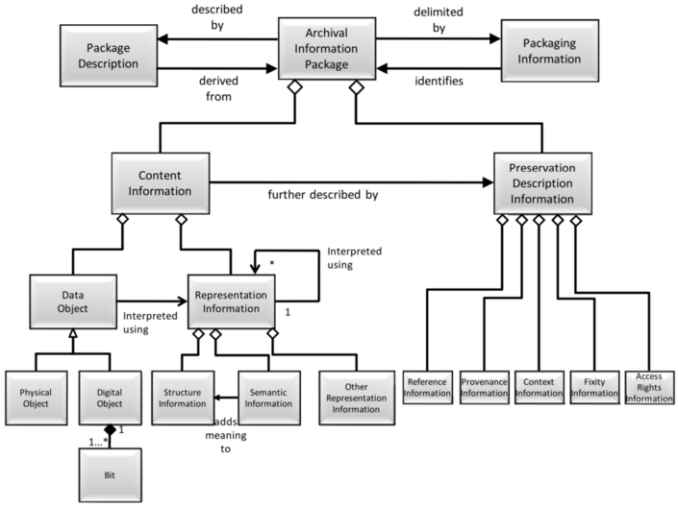


Figure 1. Schematic overview of the components of an archive information package [3]

The second step involved mapping the "metadata", i.e. all the logical classes referred to as "information" in Figure 1. This involved mapping the existing artifacts, that can be constructed from existing documents and files to the various types of information classes. Other relevant metadata may exist and will be added in later extensions of the concept.

3. Results

Based on existing experience and internal SOPs, we selected 22 types of clinical documents that were assessed as "possibly relevant". The assumed relevance resulted from demands from previous data transfers. The effort required to make them available as a published asset was estimated to be low in the majority of cases, since they already exist. However, it is questionable whether the information contained is sufficient for

further processing (example Data Sharing Plan: intention to share "yes" is too unspecific). Internal processes, names of natural persons within documents, preliminary study results were considered to be significantly more problematic for publication. Even if accessibility can be restricted by defining rights and user roles, only low-problem resources should be published first. The remaining document types were examined for their expected usefulness for third party interpretation of the study. Ultimately, the following document types should be captured in an AIP (if available): a) data dictionary and/or (annotated) case report forms, b) schedule of visits and assessments, c) data quality rules, d) algorithms or scripts for transformation or evaluation, e) records in clinical registries, f) research data management plan, g) scientific publications and supplements, h) datasets.

The actual dataset (the OAIS Digital Object) is a special case. Although there cannot be a low risk here due to its nature, since it contains sensitive personal health data, its availability is the purpose of the project. The second step involved mapping the "metadata", i.e. all the logical classes referred to as "information" in Figure 1. This involved mapping the existing artifacts contained in the document (see Table 1). However, in practice, metadata is distributed in different source assets, which requires transformations from the Submission Information Package (SIP) to the AIP. Some metadata is present multiple times and in different granularity, some not at all. An example of this is an export of data in CDISC-ODM format with rich metadata compared to a CSV export without metadata.

Table 1. Exemplary mapping of metadata elements corresponding to types of information classes. Container *Content Information, Representation Information* and *Preservation Description Information* omitted.

Information classes (AIP)	Metadata elements relevant for clinical trials
Structure Information	• Datatype, format, precision of variables in the dataset
Semantic Information	• Representation of the logical structure of the dataset, e.g.forms that data elements belonged to or events there were collected • Annotations with concepts from medical terminologies like CDISC CDASH/SDTM, LOINC or SNOMED CT
Other Information	• Standards the data object depends on: XML, CDISC ODM, UFT8
Reference Information	• Authors, name, title, version, … • Identifiers like DOIs or internal LHA-IDs
Provenance Information	• Source systems of the dataset and others from the set of 19 data elements as defined in [6] • Audit trail (change log to dataset)
Context Information	• Funding, related projects • References to other parts for composite datasets • Documents helping to understand the experiment like publications or SOPs
Fixity Information	• Electronic signatures from investigators (from CDISC ODM) • Certificates or cryptographic keys for privacy-preserving distributed analysis (DataSHIELD in planning) or pseudonymization algorithms of patient-identifying data like names or addresses
Access Rights Information	• Rights and roles • Licence information • Intellectual property contracts such as publication moratoria or data use agreements
Packaging Information	• File structure in ISA/SEEK
Package Description	• Study acronym, conditions observed, sample size and 8 others from the NFDI4Health metadata schema [7]

An obvious challenge is metadata that is not included in the SIP but is important for data sharing, such as licenses or data usage agreements.

4. Discussion and Conclusions

Many projects constantly generate data that they offer in repositories for subsequent use. Once established, the question of optimized storage of data arises to support subsequent research questions and ensure reproducibility. The concept presented here is just currently being implemented in the Leipzig Health Atlas (LHA) [8]. One limitation is that the concept has not yet been evaluated by external partners and projects. Furthermore, no harmonized vocabularies exist for parts of the metadata, e.g., on informed consents. It will be expectedly challenging to get high quality metadata from data providers who need to do manual research on it, e.g., providence information. In the future, the LHA will serve as a template for "Local Data Hubs (LDH)" of NFDI4Health [9], which requires standards-compliant, machine-readable vocabularies and interfaces [10]. From a clinical perspective, it will be important to set the bar low for discoverability and access.

The OAIS standard is widely used and is currently being further elaborated [11]. OAIS is the dominant standard due to its age and influence, most alternatives are conceptually based and address additional facets such as trustworthiness (ISO 16363) and planning (Digital Curation Centre Curation Lifecycle Model) [12]. For clinical trials, there is an EMA "guideline on the content, management and archiving of the clinical trial master file", which refers solely to archiving and covers organizational and financial issues.

Acknowledgements: Research was supported by DFG grant WI 1605/10-2, SA 1009/3-2 and BMBF grant 031L0026.

References

[1] Semler SC, Wissing F, Heyder R (2018) German Medical Informatics Initiative. Methods Inf Med 57, e50-e56.
[2] Fluck J (2021) National Research Data Infrastructure for Personal Health Data (NFDI4Health) doi: 10.4126/FRL01-006430386.
[3] ISO 14721:2012: Space data and information transfer systems — Open archival information system (OAIS) — Reference model. International Organization for Standardization, Geneva, Switzerland. https://www.iso.org/standard/57284.html
[4] Schrimpf S (2014) Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven, Beuth, Berlin.isbn: 9783410239543
[5] Sinaci AA, Núñez-Benjumea FJ, Gencturk Met.al. (2020) From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. Methods Inf Med 59, e21-e32.
[6] Henke C, Graf L, Kuntz AS et.al. (2022) The Way Data Flows: Current Provenance Options in Collaborative Research. 67. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS).
[7] Klopfenstein SAI, Golebiewski M, Schmidt CO et.al. (2022) NFDI4Health Task Force COVID-19 Metadata schema V2_0. doi: 10.4126/FRL01-006431357
[8] Meineke FA, Löbe M, Stäubert S (2018) Introducing Technical Aspects of Research Data Management in the Leipzig Health Atlas. Studies in health technology and informatics 247, 426–430.
[9] Kirsten T, Meineke F, Löffler-Wirth H et.al. (2022) The Leipzig Health Atlas - An open platform to present, archive and share bio-medical data, analyses and models online. Methods Inf Med.
[10] Löbe M, Ulrich H, Beger C et.al. (2022) Improving Findability of Digital Assets in Research Data Repositories Using the W3C DCAT Vocabulary. Studies in health technology and inf. 290, 61–65.
[11] David Giaretta, John Garrett, Mark Conrad, Eld Zierau, Terry Longstreth, John Steven Hughes, Matthias Hemmje, Felix Engel (2022) OAIS Version 3 Draft Updates, Open Science Framework. doi: 10.17605/OSF.IO/93AED
[12] Higgins S (2008) The DCC Curation Lifecycle Model. International Journal of Digital Curation 3, 134–140.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230248

Model and Strategy for Predicting and Discovering Drug-Drug Interactions

Abdelmalek MOUAZER[a,1], Nada Boudegzdame[a], Karima SEDKI[a], Rosy TSOPRA[b,c,d] and Jean-Baptiste LAMY[a]
[a] *Université Sorbonne Paris Nord, LIMICS, INSERM, F-93000, Bobigny, France*
[b] *INSERM, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, F-75006 Paris, France*
[c] *HeKA, INRIA Paris, France*
[d] *Department of Medical Informatics, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France*

Abstract. Taking several medications at the same time is an increasingly common phenomenon in our society. The combination of drugs is certainly not without risk of potentially dangerous interactions. Taking into account all possible interactions is a very complex task as it is not yet known what all possible interactions between drugs and their types are. Machine learning based models have been developed to help with this task. However, the output of these models is not structured enough to be integrated in a clinical reasoning process on interactions. In this work, we propose a clinically relevant and technically feasible model and strategy for drug interactions.

Keywords. Drug Drug interaction, Modeling, Machine learning, Decision support Systems

1. Introduction

The aging of the population, the onset of many chronic diseases, and the resistance to treatment often necessitate the use of multiple medications. Drug combinations can be beneficial and interesting, but they can also lead to problematic situations for the patient. Indeed, if we take into account the different interactions, we can take advantage of the synergistic effects between several molecules to propose an effective treatment, on the other hand, if we do not take into account all the interactions, this can lead the patient to dangerous or fatal iatrogenic accidents. To take into account drug interactions, clinicians usually rely on manually structured databases based on data from the literature. However, the absence of an interaction in the databases does not mean that it does not really exist. In fact, the absence of evidence does not mean the evidence of absence. A search on PubMed with the query "drug interactions" returns 336,256 results without counting the interactions and articles duplication. On DrugBank [1], we can count 15,229 substances[2]. The number of possible interactions

[1]Corresponding Author: Abdelmalek MOUAZER, Université Sorbonne Paris Nord, LIMICS, INSERM, F-93000, Bobigny, France; E-mail: malikmouazer@gmail.com.

[2] DrugBank: https://go.drugbank.com/about

between all these substances is the set of combinations C (2, 15229) which gives a total of 115,953,606. In addition, there can be several interactions between the same pair of substances.

To address the challenge of conducting clinical trials on all possible interactions, alternative solutions based on machine learning algorithms have emerged. To predict unknown drug interactions [2]. However, most existing machine learning models predict only whether two drugs may interact, without specifying the type of interaction. While there are proposed models that predict the description of the interactions [3,4], the output data is often in the form of unstructured text.

To overcome this limitation, we propose a structured and simplified output data model for drug interaction prediction that is both machine-feasible and contains sufficient information to compare automatically the predictions with structured databases. This will enable us to confront the predictions with real-life data and use them in a decision support system for prescription analysis.

2. Methods

2.1. Formal model proposal

To define and create the model proposed in this work, we relied on the recommendations of the minimal clinically relevant information to be filled in for each interaction in clinical practice [5], the model of interactions generally found in databases on drugs (Drugbank for example), as well as the output formats of some machine learning models that predict the type of interactions [3,4].

It was also deemed necessary to distinguish between pharmacodynamic (PD) interactions that have consequences on the effects of the drug on the body and pharmacokinetic (PK) interactions that have consequences on the effect of the body on the drugs.

2.2. Strategy and reasoning

The strategy used in this work consists in linking the different interactions that may exist in a set of substances in order to detect interactions between more than 2 drugs that may contribute to the same effects according to the following:

- A PD interaction between n drug: includes all drugs that are involved in the set of 2 by 2 interactions that share the same effect.
- A PK interaction between n drug: includes all drugs that are involved in the set of 2 by 2 interactions that modify the Bio-availability of a given substance.

2.3. Proof of concept

To show the feasibility and usefulness of our model and strategy, we have taken up the work of Yan *et al* [4] who developed a model and a database of interaction types obtained by machine learning techniques. This data is freely accessible on github[3]. The

[3]https://github.com/YifanDengWHU/DDIMDL/

predicted interaction types are in unstructured text format. It contains information on the action of the interaction (increase or decrease) but no information about the "victim drug" and the "perpetrator drug". There is also no distinction between PD & PK interactions. We developed a program that reclassifies the interactions present in this database in the proposed data format. We then manually validated the automatic classification.
Finally, we applied the strategy described above to generate two tables summarizing all the interactions present in a set of arbitrarily chosen drugs.

3. Results

3.1. Structured and simplified interactions model

We propose the following attributes to describe an interaction:
Substance: The active ingredients involved in the interactions.
Description: Narrative text explaining the interaction in a textual way.
Type: Description of the meaning of the interaction. Is it "Drug1" that acts on "Drug2", the reverse or the addition of the effects of the two substances. It is a question of determining the "victim drug" and the "perpetrator drug".
Action: Qualifies the nature of the interaction in relation to the target property. Is it an increase or a decrease of this property.
Property: We classify the properties of drugs targeted by an interaction with a pharmacological logic. We thus suggest two main types of interactions:
Pharmacodynamic interaction: In this type of interaction, the nature of the target property is a modification of an effect of the drug on the body. This effect can be the therapeutic effect or an adverse effect. This interaction therefore targets a single effect.
Pharmacokinetic interaction: In this type of interaction, the nature of the target property is a modification of the bio-availability of the active ingredient. This interaction can occur in several phases of the pharmaceutical active substances, namely: Absorption, Distribution, Metabolism, Elimination. Depending on the phase of occurrence of the interaction, the consequence on the bio-availability differs. For example, a decrease in absorption decreases bio-availability. On the contrary, a decrease in elimination increases bio-availability. Finally, this interaction can concern all the effects of the "victim-drug". Indeed, if the bio-availability of an active ingredient is decreased, there is a decrease in adverse effects but also a risk of therapeutic failure. Conversely, there is a risk of toxicity and of the appearance and accentuation of undesirable effects.
Finally, in this model, we consider that the same pair of substances can be the object of several interactions (Figure 1).

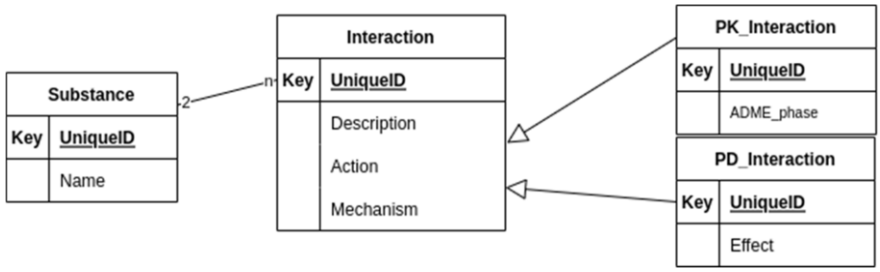


Figure 1. Structured and simplified model for the prediction of drug-drug interactions and their type ADME_phase : Acronym for (Absorption, Distribution, Metabolism or Elimination phases); PK = pharmacokinetic, PD: pharmacodynamic, Key : Primary key.

3.2. Application on machine learning predictions

In Yan *et al* [4] study, the authors developed a method to predict and classify interactions. We built upon their work by incorporating our proposed model (Figure 1) to structure the various types of interactions. The original work contains 37264 interactions divided into 65 types. Each interaction was manually annotated and assigned a structured type based on our proposed model. This step took 5 hours since the number of distinct types were small. This resulted in a comprehensive map that we used to automate the process of re-structuring the interactions.
Using this map, we classified the 37,264 interactions into 19,022 PK, which consisted of 30 A, 0 D, 10799 M, 161 E, and 8032 bio-availability interactions, without specifying the ADME phase. Additionally, we classified 18242 PD interactions, which included 1477 therapeutic effects (without specifying the effect), 9496 adverse effects (without specifying the effect), and 7269 particular effects.

3.3. Database and proposal for decision support

The database generated in this work is available in a freely accessible github repository[4]. A python program that can be found in the same repository makes it possible to generate decision support tables for researchers or developers of decision support systems.

4. Discussion

With the proposed output data model for drug interaction prediction, it becomes possible to take into account all the interactions involved in a set of substances. This allows us to consider interactions between multiple drugs, including 2-to-2 interactions that share the same effect or PK interactions that modify the bio-availability of a drug that fits or does not fit into PD interactions. We believe that this method can be integrated into a comprehensive process of monitoring and discovering new clinically significant drug interactions.
By aligning the model with MEDDRA terminology, we can query pharmacovigilance databases and health data warehouses to attribute adverse events to one or more drug interactions. This may lead to the discovery of new interactions based on real-life data

[4]https://github.com/MalikMouazer/ddireasoninglite

or raise the need to conduct clinical trials to support the evidence of a predicted interaction with machine learning algorithms.
Our proposed model does not take into account certain variables that are relevant in clinical practice, such as the severity of the interaction and recommended actions. Although this information may be of interest, we have not found data based on machine learning predictions that incorporate these variables.

5. Conclusions

The data model we propose is a compromise between the models of structured databases, which are often too complex, and the models obtained by machine learning techniques which are not structured enough. This work demonstrates the technical feasibility and clinical utility of our approach.
Our proposed model for drug-drug interactions (DDIs) provides a simple and structured way of representing interactions, which can be beneficial for training machine learning (ML) algorithms.
In perspective, we plan to fully automate the process of predicting and structuring interactions to adapt to the proposed model.

Acknowledgments

This work was partly funded by the French Research Agency (ANR) through the ABiMed project [grant number ANR-20-CE19-0017-02].

References

[1] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D1074-D1082. doi: 10.1093/nar/gkx1037. PMID: 29126136; PMCID: PMC5753335.

[2] Han K, Cao P, Wang Y, Xie F, Ma J, Yu M, Wang J, Xu Y, Zhang Y, Wan J. A Review of Approaches for Predicting Drug-Drug Interactions Based on Machine Learning. Front Pharmacol. 2022 Jan 28;12:814858. doi: 10.3389/fphar.2021.814858. PMID: 35153767; PMCID: PMC8835726.

[3] Feng YH, Zhang SW, Shi JY. DPDDI: a deep predictor for drug-drug interactions. BMC Bioinformatics. 2020 Sep 24;21(1):419. doi: 10.1186/s12859-020-03724-x. PMID: 32972364; PMCID: PMC7513481..

[4] Yan XY, Yin PW, Wu XM, Han JX. Prediction of the Drug-Drug Interaction Types with the Unified Embedding Features from Drug Similarity Networks. Front Pharmacol. 2021 Dec 20;12:794205. doi: 10.3389/fphar.2021.794205. PMID: 34987405; PMCID: PMC8721167.

[5] Hochheiser H, Jing X, Garcia EA, Ayvaz S, Sahay R, Dumontier M, Banda JM, Beyan O, Brochhausen M, Draper E, Habiel S, Hassanzadeh O, Herrero-Zazo M, Hocum B, Horn J, LeBaron B, Malone DC, Nytrø Ø, Reese T, Romagnoli K, Schneider J, Zhang LY, Boyce RD. A Minimal Information Model for Potential Drug-Drug Interactions. Front Pharmacol. 2021 Mar 8;11:608068. doi: 10.3389/fphar.2020.608068. PMID: 33762928; PMCID: PMC7982727.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230249

Building an Ontology for Traditional Medicine by Comparing Traditional Medicine Information of Chapter 26 of ICD-11 with SNOMED CT

Tom LUNDHOLM[a] and Stefano BONACINA[b,1]
[a]*Joint Master's Programme in Health Informatics, Karolinska Institutet, Solna, Sweden*
[b]*Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Solna, Sweden*
ORCiD ID: Stefano Bonacina https://orcid.org/0000-0002-5717-0647

Abstract. Chapter 26 of the 11th revision of the International Classification of Diseases (ICD-11-CH26) has introduced Traditional Medicine knowledge for use and integration with Western Medicine. Traditional Medicine is the use of beliefs, theories, and experiences to provide healing and care. The amount of information on Traditional Medicine in Systematized Nomenclature of Medicine – Clinical Terms (SCT), the world's most comprehensive health terminology, is unclear. The purpose of this study is to address this unclarity and investigate to which extent the concepts of ICD-11-CH26 can be found in SCT. If a concept from ICD-11-CH26 has a corresponding, or similar, concept in SCT, the hierarchical structure of the concepts has been compared. Then, an ontology of Traditional Chinese Medicine using the concepts of SCT will be developed.

Keywords. SNOMED CT, ICD-11, Medicine, traditional, Biomedical Ontologies

1. Introduction

Western Medicine (WM) refers to the use and application of community, science, and scientific results to create evidence-based healing, treatments, medications, and care [1–4]. This includes applying different scientific disciplines to understand the human body on all levels [1–3]. In WM, a popular belief is that the interaction between molecules is the creator of all living matter and that all matter consists of their own components [2,4].

One of the commonly used Traditional Medicine (TM) practices in the WHO member states is Traditional Chinese Medicine (TCM) [5]. TCM means TM practices that originate from ancient Chinese traditions, characterized by using experience-based and belief-based theories and treatments to create balance within the human body and in relation to its external environment [3,5]. It can be practiced for treatment and diagnosis without the need to scientifically understand the pathogenesis or cause of a health state [5]. Many of the practices and principles used in TCM derive from the view that natural

[1] Corresponding Author: Stefano Bonacina, Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, 17177 Solna, Sweden; E-mail: stefano.bonacina@ki.se.

phenomena can be categorized in two interdependent and opposite aspects of nature [5]: Yin and Yang. In TCM, an organism's material aspect is referred to as Yin and the function of the organism as Yang [5]. Body harmony is created when Yin and Yang are in balance [5]. TCM is also built on the five-element theory, which describes that the universe consists of the five elements: earth, fire, metal, water, and wood [5,6]. These describe the relationship of the human body and the external environment, and the internal interactions of the human organs [5,6].

1.1. Medical Documentation

Documentation is an important aspect of any medical approach [7]. For WM, medical documentation (MD) should contain relevant information and a professional language [8,9]. MD can aid in creating standard approaches to patient care, serve as records and be used in future work as references [7]. It is important that MD fulfill the purpose it has been designed for [7,8]. For information within the MD to be used efficiently, it needs to be provided in an appropriate format, when it is necessary and where it is needed [8]. One of the approaches to perform MD is by using medical terminologies and ontologies [9,10]. MD of TM differs from the MD of WM because much of the TM knowledge is native to the region where it is practiced and handed down in the local language through writing or oral communication [7,11,12].

1.2. Problem description and research aims

MD of TM does not always occur or follow a standardized procedure [11,12]. For much of the documented TM knowledge, the information of it is often scattered and presented using non-standardized formats [9]. Not following a standardized procedure for documentation of TM knowledge may hinder the preservation of it [13]. Standardization of TM documentation can provide many societal benefits, particularly to the societal parts related to health [13]. It could promote education, improve the safety and quality of healthcare as well as allow for evidence-based research on TM and its effects [13].

This study carries two aims: 1) To provide an overview of the concepts of ICD-11-CH26 that are available in SCT; 2) To provide a simplified way of using TM concepts in WM practices by beginning the process of building an Ontology of Traditional Medicine using the concepts available in SCT.

2. Methods

The study was conducted from January 2022 to May 2022, using the ICD-11 browser version 02/2022, SCT browser version 28-02-2022, Cambridge dictionary, Medical Subject Headings controlled vocabulary thesaurus as well as the ontology development environment Protégé. The data collection consisted of having the information from ICD-11-CH26 and SCT available for use and referencing. The data has been analyzed using qualitative content analysis, which means analyzing the context and the participant of categories that are related [14]. The purpose of this analysis method is to identify any similarities or differences that may exist [14].

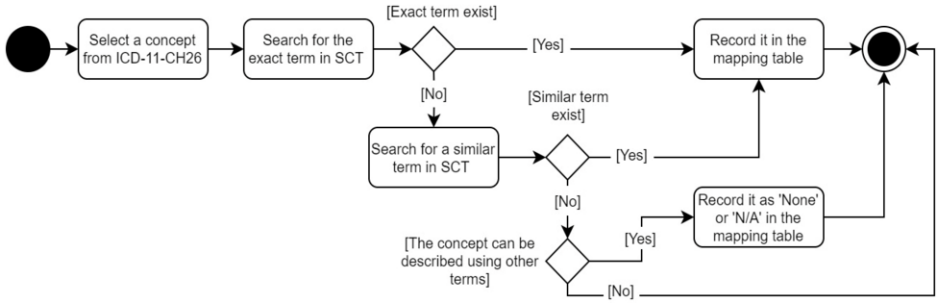


Figure 1. The term mapping method developed in this study, shown as Unified Modelling Language (UML) Activity Diagram (AD).

To employ the qualitative content analysis, the context was represented by the subject area of TM, and the participants were represented by ICD-11-CH26 and SCT concepts. A methodology based on the term mapping process described by Rodrigues et. al. [15] was developed for this study and has been described in Figure 1. An exact term was defined as where the wording of the terms is the same in both ICD-11-CH26 and SCT. A similar term was defined as using a different name, synonym(s) or phrasing that carries the same meaning as the original term.

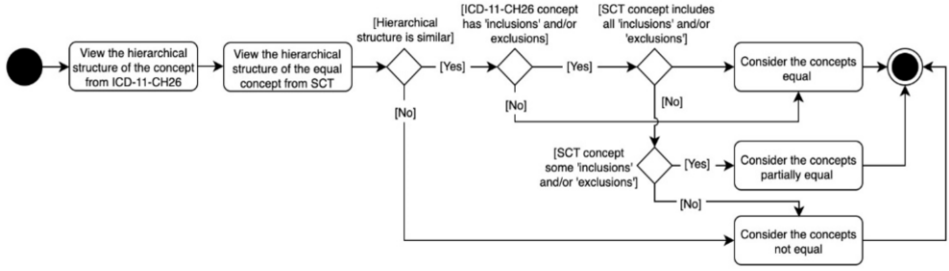


Figure 2. The semantic and hierarchical assessment method developed in this study, shown as UML AD.

For the concepts from ICD-11 that had an equal or similar concept in SCT, the semantic network and hierarchical structure of those concepts were assessed. Figure 2 gives a detailed description of the semantic and hierarchical assessment done in the study.

The ontology of this study, the Ontology of Traditional Medicine (OTM), has been developed based on a SNOMED CT Clinical Finding ontology developed by Tania Tudorache in 2014 [16], henceforth referred to as the 'Tudorache ontology' (TO). This ontology was selected due to the following reasons: 1 - To re-use existing SCT knowledge instead of creating a new and possibly overlapping knowledge. This allowed the development of the OTM to prioritize the representation of TM knowledge; 2 - The concepts identified through the term matching and the hierarchical assessment existed in the SCT domain 'Clinical Finding', either as a '(finding)' or as a '(disorder)'. Because of this, only using a Clinical Finding ontology, instead of an ontology representing the entire SCT knowledge, allowed the development of the OTM to only consider the

relevant Clinical Finding concepts; 3 – The TO was not up to date with the SCT version 28-02-2022, but contained many of the concepts identified during the data analysis.

3. Results

3.1. Disorders of ICD-11-CH26 in SNOMED CT

Of the 218 disorders represented in ICD-11-CH26, 84 had an equal or similarly named main, inclusion or exclusion concept in SCT. The list of 84 concepts and 218 disorder codes of ICD-11-CH26 can be provided upon request.
61 of the 84 concepts were considered equal to their identified SCT counterparts. These accounted for 27 percent of the total number of disorder concepts in ICD-11-CH26. For these 61 concepts, the hierarchical structures were similar and the main concepts, as well as any inclusions or exclusions, were fully representable with SCT concepts.
17 of the 84 ICD-11-CH26 concepts were considered partially equal to their identified SCT counterparts. In these cases, either the main concept was representable in SCT, but any number of inclusions or exclusions were not, or vice versa.
Six of the ICD-11-CH26 concepts were considered not equal to their equally or similarly named SCT counterparts. These concepts were not considered equal due to the ICD-11-CH26 concept and its equally or similarly named SCT concept either had entirely different hierarchical structures or an identified difficulty to represent a single ICD-11-CH26 concept with a single SCT concept.

3.2. Patterns of ICD-11-CH26 in SNOMED CT

Of the 260 patterns in ICD-11-CH26, one pattern concept had a considered equal concept in SCT. The ICD-11-CH26 pattern concept was "SE91 Qi stagnation pattern (TM1)" and the SCT equal concept entity was named "370533008 | Stagnation of chi (finding) |". The representable pattern corresponded to 3 percent of the total number of patterns in ICD-11-CH26. The other 259 ICD-11-CH26 pattern concepts were not considered fully or partially representable using SCT concepts.

3.3. Ontology of Traditional Medicine (OTM)

The OTM was designed as an adaptation of the TO. In total, there were 62 main ICD-11-CH26 concepts considered equal to their SCT counterpart. 61 of these were disorder concepts and one was a pattern concept. There were 65 SCT concepts used in this study to equally represent the corresponding ICD-11-CH26 concepts. These have been modeled and represented in the OTM. Of these, 60 existed in the TO. For 20 of these concepts, their hierarchical structure was updated according to the SCT version used. The OTM, as well as a graphical representation of it, has been uploaded in a publicly available GitHub repository [17], created and owned by the principal investigator, under the GNU General Public License v3. The TO originally had 106621 classes represented in the Clinical Finding SCT category. The total number of classes in the ontology is 106642.

4. Discussion and Conclusions

This study has compared the information of TM available in ICD-11-CH26 with the TM information available in SCT and investigated how an ontology of Traditional Medicine can be built upon this information. The results show that only a minority of the TM disorders and patterns information in ICD-11-CH26 are representable in SCT. Particularly, SCT currently does not contain much information related to the five-element theory which is a cornerstone of TCM, which makes the representation of the pattern concepts of ICD-11-CH26 difficult. The OTM that has been developed shows how some of the TM concepts of ICD-11-CH26 can be represented in SCT. Because of the design of the OTM, reusability and future development of TM knowledge is encouraged. Future work consists of analyzing how all TM knowledge of ICD-11-CH26 can be represented in SCT and a mapping between the TM concepts and Herbal Anatomical Therapeutic Chemical (ATC) codes, to better model the traditional medicine practice.

References

[1] Silvano G. A brief history of Western medicine. J Tradit Chin Med Sci. 2021;8:S10–6, doi:10.1016/j.jtcms.2020.06.002
[2] Attena F. Limitations of Western Medicine and Models of Integration Between Medical Systems. J Altern Complement Med. 2016;22(5):343–8, doi: 10.1089/acm.2015.0381.
[3] Tsuei JJ. Eastern and Western Approaches to Medicine. West J Med. 1978;128(6):551–7.
[4] Huang CD, Liao KC, Chung FT, Tseng HM, Fang JT, Lii SC, et al. Different perceptions of narrative medicine between Western and Chinese medicine students. BMC Med Educ. 2017;17(1):85, doi: 10.1186/s12909-017-0925-0
[5] Tang JL, Liu BY, Ma KW. Traditional Chinese medicine. The Lancet. 2008;372(9654):1938–40 , doi: 10.1016/S0140-6736(08)61354-9
[6] Lao L, Xu L, Xu S. Traditional Chinese Medicine. In: Längler A, Mansky PJ, Seifert G, editors. Integrative Pediatric Oncology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 125–35 , doi: 10.1007/978-3-642-04201-0_9
[7] Gutheil TG. Fundamentals of medical record documentation. Psychiatry (Edgmont). 2004;1(3):26-28.
[8] Knaup P, Benning NH, Seitz MW, Eisenmann U. eHealth and Clinical Documentation Systems. Stud Health Technol Inform. 2020;274:174-188. doi: 10.3233/SHTI200676.
[9] SNOMED - 5-Step Briefing [Internet]. SNOMED. [cited 2023 Jan 10]. Available from: https://www.snomed.org/snomed-ct/five-step-briefing
[10] World Health Organization. ICD-11 Fact Sheet [Internet]. ICD-11. [cited 2023 Jan 10]. Available from: https://icd.who.int/en/docs/icd11factsheet_en.pdf
[11] World Health Organization. Part 3 - New in ICD-11, 3.10 Traditional Medicine conditions - Module 1 (TM1) [Internet]. ICD-11 Reference Guide. 2022 [cited 2023 Jan 10]. Available from: https://icdcdn.who.int/icd11referenceguide/en/html/index.html#part-3-traditional-medicine-conditions-module-1-tm1
[12] Shojaee-Mend H, Ayatollahi H, Abdolahadi A. Development and Evaluation of Ontologies in Traditional Medicine: A Review Study. Methods Inf Med. 2019 Dec;58(06):194–204, doi: 10.1055/s-0040-1702236
[13] Rodrigues JM, Schulz S, Rector A, et al. ICD-11 and SNOMED CT Common Ontology: circulatory system. Stud Health Technol Inform. 2014;205:1043-1047.
[14] Kleinheksel AJ, Rockich-Winston N, Tawfik H, Wyatt TR. Demystifying Content Analysis. Am J Pharm Educ. 2020 Jan;84(1):7113, doi: 10.5688/ajpe7113..
[15] Rodrigues JM, Schulz S, Rector A, et al. Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. Stud Health Technol Inform. 2013;192:343-346.
[16] Tudorache T. SNOMED Clinical Finding [Internet]. BioPortal. 2014 [cited 2023 Jan 10]. Available from: https://bioportal.bioontology.org/ontologies/SNOMED_CF/?p=summary
[17] Lundholm T. Ontology-of-Traditional-Chinese-Medicine [Internet]. GitHub. 2022 [cited 2023 Jan 10]. Available from: https://github.com/TomLundholm/Ontology-of-Traditional-Chinese-Medicine

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230250

Development of Verified Innovation Process for Healthcare Solutions (VIPHS): A Stepwise Model for Digital Health

Eunji LEE[a,1], Bengt Arne SJÖQVIST[a], Magnus Andersson HAGIWARA[b], Hanna Maurin SÖDERHOLM[c] and Stefan CANDEFJORD[a]
[a] *Chalmers University of Technology, Gothenburg, Sweden*
[b] *University of Borås, Borås, Sweden*
[c] *Lindholmen Science Park AB, Gothenburg, Sweden*

Abstract. Many digital health projects often stop in the pilot or test phase. Realisation of new digital health services is often challenging due to lack of guidelines for the step-by-step roll-out and implementation of the systems when changing work processes and procedures are needed. This study describes development of the Verified Innovation Process for Healthcare Solutions (VIPHS) – a stepwise model for digital health innovation and utilisation using service design principles. A multiple case study (two cases) involving participant observation, role play, and semi-structured interviews were conducted for the model development in prehospital settings. The model might be helpful to support realisation of innovative digital health projects in a holistic, disciplined, and strategic way.

Keywords. Innovation in digital health, implementation strategy, service realisation

1. Introduction

It is often difficult to go from a project phase to an operational implementation phase. Especially the failure rate of digital transformation projects is huge. The reason why 70 percent of all digital transformations fail is a lack of discipline in defining and executing proper steps for digital transformations to take off and stay ahead [1].

When it comes to innovation in digital health, making the innovation projects successful is even more challenging due to the complexity in healthcare services. Digital technologies can support and improve delivery of health services [2]. Improvements often occur by overcoming challenges in communication due to time and place and allowing people and resources to interact in easier ways [3]. This means more people and systems can be involved in digital health. Thus, various activities can be created due to the increased number of people and systems, and their relationships. Moreover, ethical issues, healthcare policies, and the ecology of health information systems adds to the complexity, and once in operation, support, service and maintenance adds to the picture.

To make innovative digital health projects successful, there is a need for guidelines and models that can support utilisation of the projects with a holistic approach. The

[1] Corresponding Author: Eunji LEE, Hörsalsvägen 11, 41296 Gothenburg, Sweden, E-mail: eunji@chalmers.se.

existing recommendations and approaches are limited. Most of them are either focusing on interoperability issues, related to health information infrastructure, or concerning patients' accessibility to their own health data [4-7]. Two regions in Sweden reported that the most difficult thing in implementation phase of digital health projects is step-by-step roll-out of the systems [8]. There is a lack of models and guidelines that can support digital health innovation and utilisation in a holistic, disciplined, and strategic way.

Service design is a multidisciplinary, integrative, and holistic way of innovating or improving services to make them efficient for organisations and more usable, useful, and desirable for customers [9]. Service design choreographs interactions, processes, and technologies in complex systems to co-create value for relevant stakeholders [10]. The aim of this paper is to present a stepwise model for digital health innovation and utilisation following service design principles, and describe its development process.

2. Methods

A multiple case study was conducted to develop a stepwise model for digital health innovation and utilisation. Two cases were used for the development process in an iterative manner. For triangulation, participant observation, roleplay, and semi-structured interviews were used to collect data in prehospital settings in Sweden.

Two separate projects allowed to develop the VIPHS model. The development process was conducted in an iterative manner, which means that the results from the second project provided input to improve the first model from the first project. The first project, PrehospIT-Stroke, aimed to create better conditions for more efficient use of digital health within prehospital healthcare. It focused on the foundation for national harmonization of semantic and technical interoperability, with the acute stroke chain as example, aiming to improve IT support and solutions in the acute phase as well as in follow-up, business development and quality assurance at both local and national level. The second project, ViPHS (video support in the prehospital stroke chain) aimed to evaluate whether a collaborative assessment using video involving a neurologist could be effective to support the transportation decision for potential LVO (large vessel occlusion) patients, especially where there is a risk for considerable transport time to a thrombectomy facility due to geographical distance and/or first stop at local hospital.

Six principles of service design [11] were applied in the two projects. **Table 1** shows the principles, their meanings, and how they were applied in our studies' contexts.

Table 1. Principles of service design and application to our studies

Principles	Meanings [11]	Application to our studies
Human-centred	Considering the experience of all the people affected by the service	Taking into account the experiences of all involved people in the care process
Collaborative	Engaging all the stakeholders in the service design process	Involving all the stakeholders in the development process
Iterative	Iterating exploratory, adaptive, and experimental approach towards implementation	Developing and improving the service through several iterations
Sequential	Visualising and orchestrating the service as a sequence of interrelated actions	Visualising the care processes
Real	Researching needs in reality; prototyping ideas in reality and evidencing intangible values as physical or digital reality	Testing the service with increasing realism of the test settings

Holistic	Sustainably addressing the needs of all stakeholders through the entire service	Conducting care process analysis and process mapping of the entire service using a holistic approach in the initial stage

The first project, PrehospIT-Stroke, carried out in the following three phases:

1. Detailed care process analysis, identification of critical decision-making points, inventory analysis, and recommendation regarding technical and semantic interoperability,
2. Tests in a lab environment of technical and semantic recommendation (Connectathon [12]),
3. Simulated full-scale operational field test in an ambulance.

The first phases included process mapping of the stroke chain, inventory of process support, standards and de-facto standards around semantics and interoperability, analysis, and preparation of a first recommendation for harmonisation. In the second phase, the system structure was tested in the PrehospIT Connectathon in a lab environment, where the project partners' proprietary solutions were tested technically together in a simulated care process utilizing the proposed interoperability standards and methods. This test showed that the different systems could communicate with each other effectively without failures. In the last phase, the system was tested in a full-scale realistic operational simulation (roleplay), where an entire ambulance mission from dispatch to handover was simulated [13]. Eleven ambulance teams (22 ambulance clinicians) were recruited to the study. The teams were instructed to either start using a computerized decision support system (CDSS), described in [13], or to work as usual. Two representative patient cases with stroke-like symptoms were created. In one case, the patient had severe stroke symptoms, and in the other case, the patient had moderate stroke symptoms. All simulations were filmed, and all participants were interviewed after the simulations. During the semi-structured interview (average 35 minutes), questions regarding the CDSS and the system were asked; how compatible the CDSS was with their current way of working, what advantages the system brought compared to their current practice, and what problems and challenges they experienced. The interviews and films were then analysed, and the management of the patients with and without CDSS were compared.

The second project, ViPHS, was conducted in the following four phases:

1. Analysis of process and inventory and recommendation regarding technical and semantic interoperability
2. Tests in a simulated prehospital environment
3. Initial field test in ambulance
4. Field tests in a larger scale

In the first phase, the market for mobile video solutions was investigated. Different solutions were tested and evaluated in terms of image quality, function, malfunctions, and so on. An ambulance was equipped with a solution that was considered as the best alternative and tests on transmission capacity were carried out by driving the ambulance around in areas with poor cellular coverage. In the second phase, a simulated test (roleplays) was performed. Four ambulance teams (eight ambulance clinicians) and four neurologists from a regional stroke centre were recruited. The simulation was set up consisting of a video-equipped ambulance standing outside in a parking lot, an office, and a bench outside. Two patient cases were constructed inspired by real patient cases. In one case, a male patient was found by a colleague in an office with severe stroke symptoms, in the other case, a female patient was found outside with moderate stroke symptoms. After the simulations, each participant was asked questions regarding the simulation design and their experience in a semi-structured interview (average 39

minutes). In phase three, the system was tested in the field in a small pilot study. Video equipment was installed in three ambulances in an ambulance organization in Region Västra Götaland in western Sweden. The staff were trained in the system and new guidelines for the care of patients with stroke symptoms were written. For each video-equipped ambulance, a person was appointed as responsible for staff training, that guidelines were updated and implemented, and managed the collection of study protocols. The pilot study lasted one year and all other ambulances in the organization served as a control group. In phase four, the study was expanded both geographically and with a number of ambulances. Twelve ambulances throughout the region were equipped with video. In this phase the technical solution was also adopted to a video-platform supported and maintained by the region. Thereby a potential expansion of the service as well as full-scale operational support and maintenance beyond the project phase should be facilitated. This study has been going on for a year and is expected to finish in year 2023.

3. Results

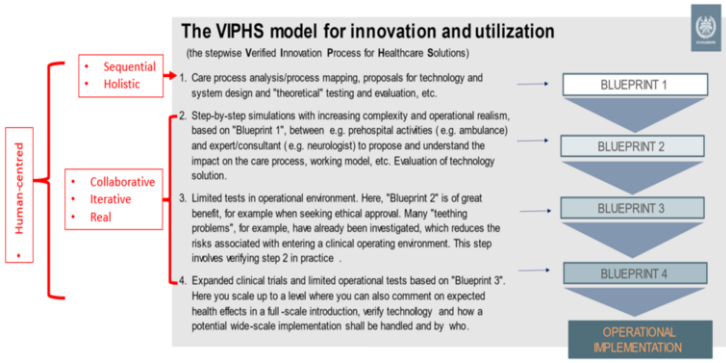


Figure 1. The VIPHS model for innovation and utilization

Figure 1 shows the stepwise model for digital health developed through the two cases. Red texts on the left side shows where the principles of service design are applied. In the VIPHS model, each step leads to a "Blueprint" which is a document with results obtained and input and recommendations to the next step. The model can also be used partially, not including all the four steps. If the project aims to define a prototype, it can stop after blueprint 2. If the project's aim is a clinical validation (e.g., operational feasibility study or clinical proof of concept), it can stop after blueprint 3. The VIPHS model provides an overall idea on what to achieve in implementation in clinical operations and allow to restart a new project from where the previous project ended.

The PrehospIT project stopped at step 3 (blueprint 3) since the goal of the project was to demonstrate the benefits for the care process deploying standards etc., enabling multi-system interoperability and improved care processes in a realistic setting. The Connectathon was the prime source for evaluating interoperability. During the user simulations the improved care process was achieved from deploying this interoperability. The ViPHS project is still running in an extended step 4 (blueprint 4) in March 2023. The clinical tests were stopped at the end of 2022. However, the tests resulted in a suggestion of modifying and updating the technical solution before going into more general deployment across the health care region. Planning is currently underway to introduce the ViPHS system in all the region's ambulances.

4. Discussion and Conclusion

In digital health, changing systems and work processes, namely digitalisation of services in healthcare, is one of the fundamental challenges. Having models that can guide and support digital health innovation and utilisations would be helpful to solve this issue. We could develop a stepwise model for digital health innovation and utilisation by applying service design principles through empirical settings. The VIPHS model with its clearly defined steps and accompanying "blueprints" could be useful to support the implementation process of digital health projects in a holistic, disciplined, and strategic way. This can contribute to increase the success rate of realizing digital health projects.

The ViPHS model's step-by-step method provides valuable data in all steps that contribute to accomplish the next step. For example, step 2 in the ViHPS project revealed that it was necessary to standardise the communication between ambulance clinicians and the neurologist. This helped to prepare step 3. In addition, the VIPHS model's stepwise approach made it easier to document and communicate our findings and recommendations to people outside the project group.

The VIPHS model was developed in a specific clinical setting, the prehospital stroke care process, which is a limitation of our study. We plan to further improve this model through more case studies in different clinical settings like trauma and fall patient care processes that use advanced technology like artificial intelligence, voice recognition, etc.

References

[1] Saldanha T. Why digital transformations fail: The surprising disciplines of how to take off and stay ahead. Berrett-Koehler Publishers; 2019.
[2] Eysenbach G. What is e-health?, J. Med. Internet Res. 2001;3:e833, doi: 10.2196/jmir.3.2.e20
[3] Mitchell J. Increasing the cost-effectiveness of telemedicine by embracing e-health, J. Telemed. Telecare. 2000;6:16–19, doi: 10.1258/1357633001934500
[4] World Health Organization. From innovation to implementation eHealth in the WHO European Region. World Health Organization. Regional Office for Europe; 2016.
[5] World Health Organization. Digital Health Platform: Building a Digital Information Infrastructure (Infostructure) for Health. World Health Organization; 2020.
[6] The European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry (COCIR). mHEALTH: COCIR recommendations for an effective development. COCIR; 2020.
[7] The European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry (COCIR). mHEALTH: Joint recommendations for an effective development. COCIR; 2021.
[8] The Government and the Swedish Association of Local Authorities and Regions. Vision for eHealth 2025 Follow-up 2019, 2020.
[9] Moritz S. Service design: practical access to an evolving field. Köln International School of Design; 2005.
[10] Mager B. Meet Birgit Mager, President of the Service Design Network. [Internet] 2023 [cited 2023 January 10] Available from: https://www.service-design-network.org/community-knowledge/meet-birgit-mager-president-of-the-service-design-network.
[11] Stickdorn M, Hormess ME, Lawrence A, and Schneider J. This is service design doing: applying service design thinking in the real world. O'Reilly Media, Inc.; 2018.
[12] Integrating the Healthcare Enterprise. Connectathon. [Internet] 2023 [cited 2023 January 15] Available from: https://www.ihe.net/participate/connectathon/.
[13] Andersson MH, et al. The effects of integrated IT support on the prehospital stroke process: Results from a realistic experiment, J. Healthc. Inform. Res. 2019;3:300–328.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230251

Insights into the FAIRness of the German Network University Medicine: A Survey

Lea MICHAELIS[a,1], Rasim Atakan POYRAZ[b], Michael Rusgoza MUZOORA[b], Kerstin GIEREND[c], Alexander BARTSCHKE[b], Christoph DIETERICH[d,e], Tim JOHANN[d], Dagmar KREFTING[f], Dagmar WALTEMATH[a,g] and Sylvia THUN[b]
[a] *Data Integration Center, University Medicine Greifswald, Germany*
[b] *Core Facility Digital Medicine & Interoperability, BIH at Charité, Berlin, Germany*
[c] *Department of Biomedical Informatics at the Center for Preventive Medicine and Digital Health, Medical Faculty Mannheim, Heidelberg University, Germany*
[d] *K. Tschira Institute for Integrative Computational Cardiology, Heidelberg, Germany*
[e] *German Centre for Cardiovascular Research (DZHK), Heidelberg, Germany*
[f] *Dpt. Of Medical Informatics, University Medical Center Göttingen, Germany*
[g] *Medical Informatics Laboratory, University Medicine Greifswald, Germany*
ORCiD ID: L. Michaelis https://orcid.org/0000-0001-9691-2677, R.A. Poyraz https://orcid.org/0000-0002-5705-7396, M.R. Muzoora https://orcid.org/0000-0002-1384-1509, K. Gierend https://orcid.org/0000-0003-0417-3454, A. Bartschke https://orcid.org/0000-0002-5849-482X, C. Dieterich https://orcid.org/0000-0001-9468-6311, T. Johann https://orcid.org/0000-0001-5792-1562, D. Krefting https://orcid.org/0000-0002-7238-5339, D. Waltemath https://orcid.org/0000-0002-5886-5563, S. Thun https://orcid.org/0000-0002-3346-6806

Abstract. The need to harness large amounts of data, possibly within a short period of time, became apparent during the Covid-19 pandemic outbreak. In 2022, the Corona Data Exchange Platform (CODEX), which had been developed within the German Network University Medicine (NUM), was extended by a number of common components, including a section on FAIR science. The FAIR principles enable research networks to evaluate how well they comply with current standards in open and reproducible science. To be more transparent, but also to guide scientists on how to improve data and software reusability, we disseminated an online survey within the NUM. Here we present the outcomes and lessons learnt.

Keywords. FAIR, Network University Medicine, Medical Informatics

1. Introduction

The determined goal of the German Network University Medicine (NUM) was to coordinate the University Hospital's COVID-19 strategies and research activities and share Covid-related data across Germany [1]. Research networks, in general, spend time and effort on implementing such strategies for data and software interoperability and sharing. This work elaborates on the current status of "FAIRness" within NUM.

[1] Corresponding Author: Dagmar Waltemath, E-mail: dagmar.waltemath@uni-greifswald.de.

2. Methods

We created a survey composed of 33 questions with over 100 possible answers [2], based on previously established questions from online FAIR frameworks and self-assessment tools. The draft questionnaire was revised with FAIR experts and the final survey thematically divided into general questions about the projects and experiences with FAIR evaluations, followed by four sections each addressing a FAIR category (Findable, Accessible, Interoperable, Reusable). Results were captured in REDCap, a secure GDPR compliant and widely used Research Electronic Data Capturing Tool. We disseminated the survey link in NUM internal meetings, the NUM newsletter and social media.

3. Results and Discussion

During the productive survey time (75 Days), only five entries from a total of three NUM projects were fully submitted even though over 100 entries had been recorded. This response rate prohibits statistically relevant statements. However, we saw that NUM projects make use of local and global identifiers for data items, have Standard Operation Procedures in place, and use common semantic interoperability standards. A specific example emphasizing on the use of FAIR principles in research software development [3] within NUM is the Anonymous Synthesizer for Health Data (ASyH, https://github.com/dieterich-lab/ASyH), an easy-to-use open-source software for creating synthetic data with machine learning models trained on real data, used and developed in the NUM CODEX+ subproject.

4. Conclusion

We notice general advertisement of the FAIR data principles in the NUM community, but the survey showed only limited implementation within the actual projects. We suggest that more information on the benefits of FAIRification should be provided and the leadership should develop a strategy for FAIR research data management including provenance information, licensing, open data repositories. Specifically, more knowledge about metadata, version control and licenses is needed.

Acknowledgements

The German Ministry of Health (BMBF) funded this project, FKZ 01KX2121. Thanks to Martin Sedlmayr and Brita Sedlmayr helping with the survey design.

References

[1] Prokosch, Hans-Ulrich, et al. "The COVID-19 data exchange platform of the German university medicine." *Challenges of Trustable AI and Added-Value on Health*. IOS Press, 2022. 674-678.
[2] Waltemath D, et al. FAIR evaluation at the NUM sites - How FAIR is NUM? Zenodo; 2022. Available from: https://doi.org/10.5281/zenodo.7299373.
[3] Barker M, et al. Introducing the FAIR Principles for research software. Scientific Data. 2022;9(1):1-6.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230252

Health-Related Content in Transformer-Based Language Models: Exploring Bias in Domain General vs. Domain Specific Training Sets

Giuseppe SAMO[a,b,1] and Caterina BONAN[c]
[a] *University of Geneva, Switzerland*
[b] *Beijing Language and Culture University, China*
[c] *University of Cambridge, UK*
ORCiD ID: Giuseppe Samo https://orcid.org/0000-0003-3449-8006,
Caterina Bonan https://orcid.org/0000-0002-4808-6865

Abstract. In this communication, we demonstrate that the bias observed in domain general training sets with health-related content is not improved in domain specific health-communication corpora, contra.

Keywords. Natural Language Processing, Health-content, Language Models, Knowledge Reproduction, Corpora, COVID-19

1. Introduction

Artificial neural language models (LMs) parse and generate complex linguistic structures across languages [1,2], and can be used for fact checking [3]. In [1], we detected syntactic bias in Transformer-based LMs using a list of myth busters on COVID-19 from parallel World Health Organization corpora. We demonstrated that when LMs are queried with sentences compared with ad hoc examples with opposite polarity, asymmetries are easily detected and quantified. Here, we explore six training sets for English and Chinese LMs to detect bias with respect to domain-specific (i.e., medical) semantico-encyclopedic knowledge, adopting and improving the dataset previously discussed in [1].[2]

2. The study

Hypothesis: In [1]'s conclusions, we tentatively attributed the observed bias to the type of training data under investigation, and hypothesised that LMs trained on domain specific datasets might perform better than those trained with domain general data.

[1] Corresponding Author: Giuseppe Samo, Giuseppe Samo, Department of Linguistics, University of Geneva, mail: Rue de Candolle 2, 1205 Geneva, Switzerland. E-mail: giuseppe.samo@unige.ch.

[2] Information on the data is available at the following link:https://github.com/samo-g/health-transformer.

Materials: The queried parallel datasets for English and Chinese are those presented in The training sets for this contribution are presented in table 1.
Methods: As in [1], the machine is presented with two sentences: (i) a TRUE statement from the corpus and (ii) a FALSE counterpart created using logical operators. Our measure is represented by the difference between the surprisal of the TRUE and the FALSE statement. The surprisal is the logarithm of the reciprocal of the output probability in a fill-mask task. In a nutshell, lower surprisal is to be expected for TRUE sentences.

Table 1. Language Models used in this paper. All references are available as hyperlinks.

Languages	Training Set Type
English	Domain general (web, wiki): BERT, BigBird
	Domain specific: HealthF (fact checking), PHS (health surveillance on social media)
Chinese	Domain general (web): Bert-base-Chinese
	Domain specific: HealthZH (medical dialogues, records, textbooks)

3. Results and Conclusions

Figure 1 displays an asymmetry for English ($F(3, 64) = 177.30362$; $p < .00001$): while the BERT general domain LM performs worse (M = 18.3, SD = 5.82) than domain specific ones, BigBird (M = -24.5, SD = 7.08) shows the least surprisal on TRUE statements, plausibly due to its architecture combining sparse and global attention.

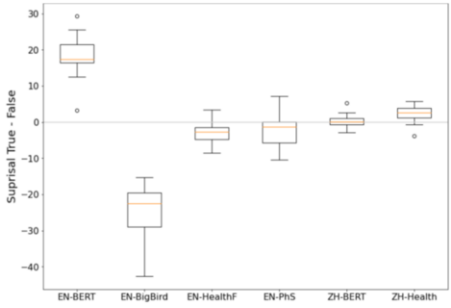


Figure 1. True-False across language models.

In Chinese, the general domain performs better (M = 0.51, SD = 2.42) than the domain-specific one (M = 1.81, SD = 2.42) ($t(34) = 2.6837$, $p < .05$). [1] provided a methodology for the detection of bias in LMs that can be adopted to fact checking and the general medical domain, and predicted lower bias with health-specific training sets. Here, we showed that domain specific training sets do not necessarily perform better than domain general ones. Future studies should explore additional datasets (e.g., online medical training, [4]) and languages.

References

[1] Samo G, Bonan C, Si F. Health-Related Content in Transformer-Based Deep Neural Network Language Models: Exploring Cross-Linguistic Syntactic Bias. Stud Health Technol Inform. 2022;295:221-225.
[2] Linzen T, Baroni M. Syntactic Structure from Deep Learning. Annual Review of Linguistics 7:1, Jan.2021;195-212.
[3] Ahmed S, Hinkelmann K, Corradini F, Development of fake news model using machine learning through natural language processing; arXiv:2201.07489, Jan. 2022.
[4] Utunen H, et al, Global Reach of an Online COVID- 19 Course in Multiple Languages on OpenWHO in the First Quarter of 2020: Analysis of Platform Use Data, J Med Internet Res. 22 (2020) e19076.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230253

Ontological Modeling of Clinical Study Forms

Jacques HILBEY[a,b,1], Amel RABOUDI [c], Marie-Odile KREBS [d] and Jean CHARLET[e,b]
[a] *Sorbonne Université, Paris, France*
[b] *Sorbonne Université, Sorbonne Paris Nord, INSERM, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé - LIMICS, Paris, France*
[c] *Fealinx, 37 rue Adam Ledoux 92400 Courbevoie, France*
[d] *Institute of Psychiatry and Neuroscience of Paris, Université de Paris, GHU Paris Psychiatrie & Neurosciences, Paris, France.*
[e] *Assistance Publique-Hôpitaux de Paris, Paris, France*
ORCiD ID: Jacques Hilbey https://orcid.org/0000-0002-3351-9840

Abstract. The use of eCRFs is now commonplace in clinical research studies. We propose here an ontological model of these forms allowing to describe them, to express their granularity and to link them to the relevant entities of the study in which they are used. It has been developed in a psychiatry project but its generality may allow a wider application.

Keywords. Biological Ontologies, Clinical Study, Form.

1. Introduction

Our proposal, retaining may elements of Bona's proposal [1] as well as proposals extending Bona's [2][3], aims at the best access to the data and its contextualization for future reuse. The proposed strategy includes maintaining the link to the original data files, ontological representation of their schema, enrichment with metadata.

2. Material and methods

The ontoPSYCARE modular ontology is being developed within the PsyCARE project[2] to allow data integration as well as semantic annotation. An event-centered foundational ontology (ontoPOF) and a data-centered core ontology (ontoDOME) provides the ontological commitment explaining the proposed representation of the assessment tools in the module dedicated to psychiatry data (ontoDOPSY).

[1] Corresponding Author: Jacques Hilbey; E-mail: jacques.hilbey@sorbonne-universite.fr.
[2] https://psy-care.fr/

3. Results

The result is a threefold model (Fig. 1) providing an overview of the assessment tool, that also allows the tool and the responses obtained from a patient to be enrolled in compositional relationships, and including the form in the entities representing its completion.

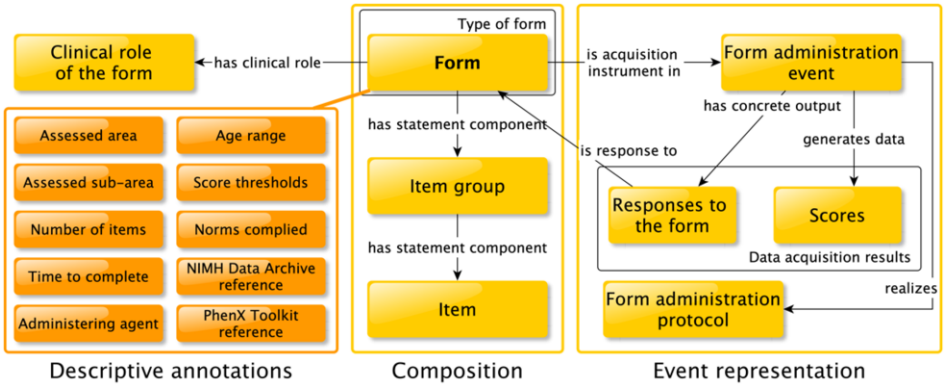


Figure 1. A threefold model (descriptive, compositional, event-related) of an assessment tool.

4. Discussion and conclusion

The model we propose here aims first at allowing the integration of data produced by a clinical research project on a data collection platform (while ensuring the traceability of data flows and treatments thanks to the BMS-LM ontology [5], which uses PROV-O), and an ontological representation centered on the form administration events is also provided ; it also allows the exploration of the assessment tools and the assessments.

5. Acknowledgments

This work has been supported by the French government's “Investissements d'Avenir” program, which is managed by the Agence Nationale de la Recherche (ANR), under the reference PsyCARE ANR-18-RHUS-0014.

References

[1] Bona J, Kohn G, Ruttenberg A. Ontology-driven patient history questionnaires. Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO 2015), vol. 1515, CEUR vol. 1515; 2015.
[2] Gonçalves, R., Tu, S., Nyulas, C. et al. An ontology-driven tool for structured data acquisition using Web forms. J Biomed Semant 8, 26 (2017). doi: 10.1186/s13326-017-0133-1
[3] Fabry P, Barton A, Ethier JF. QUESTO – An Ontology for Questionnaires. Proceedings of the Eleventh International Conference on Biomedical Ontology (ICBO 2020), vol. 2807, CEUR vol. 2807; 2020.
[4] Hilbey J, Aimé X, Charlet J. Temporal Medical Knowledge Representation Using Ontologies. Stud Health Technol Inform. 2022 May 25; 294:337-341. doi: 10.3233/SHTI220470.
[5] Raboudi A *et al.* The BMS-LM ontology for biomedical data reporting throughout the lifecycle of a research study: From data model to ontology. J Biomed Inform. 2022 Mar;127:104007.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230254

Exploring Use Cases for CovidGraph

Lea GÜTEBIER[a,1], Ron HENKEL[a], Christian NIKLAS[b] and Dagmar WALTEMATH[a]
[a] *Department of Medical Informatics, Institute for Community Medicine, University Medicine Greifswald, Germany*
[b] *Institute for Medical Informatics, University Hospital Heidelberg, Germany*

Abstract. HealthECCO is the driving force behind the COVID-19 knowledge graph spanning multiple biomedical data domains. One way to access CovidGraph is SemSpect, an interface designed for data exploration in graphs. To showcase the possibilities that arise from integrating a variety of COVID-19 related data sources over the last three years, we present three use cases from the (bio-)medical domain. **Availability:** The project is open source and freely available from: https://healthecco.org/covidgraph/. The source code and documentation are available on GitHub: https://github.com/covidgraph.

Keywords. Data integration, Exploration, Knowledge graphs, Covid-19

1. Introduction

Knowledge graphs have become a significant tool for data exploration in biomedicine over the recent years. The underlying graph model represents heterogeneous data as nodes and emphasises the relationships between individual data points. A graph model offers a high flexibility and better usability for data exploration, for example, when looking for clinical studies, exploring simulation models, or investigating disease mechanisms [1].

HealthECCO is a community of graph enthusiasts who first united forces in the wake of the COVID-19 pandemic. CovidGraph, a tool for efficient data access and exploration to support research efforts relating to COVID-19 [2], is HealthECCO's fast response to the pandemic. As of today, CovidGraph integrates a variety of data sources, e.g., scientific publications, study protocols, and simulation studies.

To exploit the value in the connected data, we explored several use cases for CovidGraph.

2. Methods and Results

The data domains represented in CovidGraph include publications, patents, clinical trials, systems biology models, and biomedical ontologies. The data is represented in a labelled property graph which structures the data as labelled nodes and edges that connect individual nodes [3]. Both nodes and edges can hold properties with more detailed information about the represented data.

[1] Corresponding Author: Lea Gütebier, Walther-Rathenau-Str. 48, Department of Medical Informatics, University Medicine Greifswald, D-17475 Germany; E-mail: lea.guetebier@uni-greifswald.de.

CovidGraph is build on a Neo4j[2] graph database engine which can be accessed through different user interfaces [2]. One of these interfaces is SemSpect, a visualisation tool for interactive exploration of knowledge graphs [4]. SemSpect is designed as a drag & drop application for querying the graph without knowledge of a programming language. The tool visualises and aggregates groups of nodes based on specific labels. The user can then explore the graph and gradually expand the visualisation by including related groups. For a more detailed view, the user selects individual nodes within a group. Therefore, SemSpect is well suited to answer scientific questions in a flexible and exploratory manner. A few examples are presented in the following use cases.

Our first use case shows how to collect decisive information about a specific gene (*PF4*) which is involved in the manifestation of thromboses as a side-effect of viral vector COVID-19 vaccinations [5]. SemSpect includes synonyms of the gene name in the exploration. The resulting graph contains molecular data such as biological entities and pathways, textual data such as publications, and patents.

In our second use case, we identify potential simulation models studying the *TCF7L2* variants as those indicate a predisposition for type 2 diabetes [6]. As of now the exact mechanisms remain elusive. With SemSpect, we can explore models that investigate the influence of *TCF7L2* variants on various pathways, thereby helping to investigate the disease on a molecular level.

Our third use case outlines how to explore health studies for clinical research by finding already exiting studies similar to a researcher's own studies of interest.

3. Discussion

The three mentioned use cases showcase how CovidGraph supports knowledge exploration and increases the value and the potential of reusing healthcare data for secondary research. We encourage readers to actively explore the knowledge graph themselves[3].

References

[1] Stothers JA, Nguyen A. Can Neo4j replace PostgreSQL in healthcare? AMIA Summits Transl Sci Proc 2020;2020:646.

[2] Gütebier L, Bleimehl T, Henkel R, Munro J, Müller S, Morgner A, et al. CovidGraph: a graph to fight COVID-19. Bioinformatics 2022;38:4843–5. https://doi.org/10.1093/bioinformatics/btac592.

[3] Angles R, Gutierrez C. Survey of graph database models. ACM Comput Surv CSUR 2008;40:1–39. https://doi.org/10.1145/1322432.1322433.

[4] Liebig T, Vialard V, Opitz M. Connecting the Dots in Million-Nodes Knowledge Graphs with SemSpect. ISWC Posters Demos Ind. Tracks, 2017, p. Vol-1963:paper587.

[5] Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. N Engl J Med 2021;384:2092–101. https://doi.org/10.1056/NEJMoa2104840.

[6] Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 2006;38:320–3. https://doi.org/10.1038/ng1732.

[2] https://neo4j.com/

[3] https://db.covidgraph.org/semspect/, username: public, password: corona

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230255

Exploring New Possibilities for Research Data Exploration Using the Example of the German Core Data Set

Felix MENZEL[a,b], Dagmar WALTEMATH[a,b] and Ron HENKEL[a,1]
[a] *Department of Medical Informatics, Institute for Community Medicine, University Medicine Greifswald, Germany*
[b] *Core Unit Data Integration Center, University Medicine Greifswald, Germany*

Abstract. The German Medical Informatics Initiative (MII) aims to increase the interoperability and reuse of clinical routine data for research purposes. One important result of the MII work is a German-wide common core data set (CDS), which is to be provided by over 31 data integration centers (DIZ) following a strict specification. One standard format for data sharing is HL7/FHIR. Locally, classical data warehouses are often in use for data storage and retrieval. We are interested to investigate the advantages of a graph database in this setting. After having transferred the MII CDS into a graph, storing it in a graph database and subsequently enriching it with accompanying meta-information, we see a great potential for more sophisticated data exploration and analysis. Here we describe the extract-transform-load process which we set up as a proof of concept to achieve the transformation and to make the common set of core data accessible as a graph.

Keywords. Core Data Set, Graph Database, Data Integration, HL7/FHIR, Data Exploration

1. Introduction

The Medical Informatics Initiative (MII) [1] has established data integration centers (DIZ) at all university hospitals in Germany. Together, the large consortium has agreed on a core data set (CDS)[2] to enable data sharing for research purposes. The core data set is divided in six modules (person, case, diagnosis, procedure, laboratory test results and medication). The agreed-upon interface for data exchange across the DIZs is HL7/FHIR [2]. However, the local data stores differ across the university hospitals, e.g., due to pre-existing IT-infrastructures.
As part of the HealthECCO[3] community effort we contributed to the development of CovidGraph [3] and thereby demonstrated the power of graph databases for fast and flexible biomedical data exploration during the COVID-19 pandemic.

[1] Corresponding Author: Ron Henkel, Walther-Rathenau-Str. 48, Department of Medical Informatics, University Medicine Greifswald, D-17475 Germany; E-mail: ron.henkel@uni-greifswald.de.
[2] CDS: https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set
[3] HealthECCO: https://healthecco.org/

This work elucidates the potential of transforming and storing the MII CDS into a graph database to facilitate easy exploration using SemSpect [4] and to enable domain-spanning queries [5] in the broad field of medical informatics.

2. Data and Methods

The development of our **e**xtract, **t**ransform, **l**oad (ETL) process started with a detailed analysis of the MII CDS. We then evaluated commonly used ETL tools [6,7]. Based on this analysis, a graph model is developed and accompanying meta-information is identified. The data is accessed via the HL7/FHIR interface, mapped to the graph model, and loaded into the graph database[4]. In a last step, accompanying meta information, related clinical studies and ontology terms, will be loaded, cross-referenced and linked to the CDS.

3. Results and Discussion

Storing and linking medical data in a graph database opens new doors for data exploration. For example, SemSpect [4] supports expansion and filtering, and automatic grouping of similar data items. Taken together, these features enable to easily traverse the graph. Visual representations can be created without detailed knowledge of the underlying data model. In addition, Neo4j Bloom[5], an application for graph exploration, offers semi-natural language queries, rule-based styling and allows to search for phrases and pattern. Results from the different exploration steps can be visualized or accessed in tabular format as well as attribute-value pairs. Future plans include the implementation of sophisticated data analysis procedures using graph algorithms. As the core data set contains medical data, this proof of concept will obey to the same data access restrictions as the original data set.

References

[1] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018;57(S 01):e50–6. http://dx.doi.org/10.3414/ME18-03-0003.
[2] Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities. JMIR Med Inform. 2021;9(7):e21929. http://dx.doi.org/10.2196/21929.
[3] Gütebier L, Bleimehl T, Henkel R, Munro J, Müller S, Morgner A, et al. CovidGraph: a graph to fight COVID-19. Bioinformatics. 2022;38(20):4843–5. http://dx.doi.org/10.1093/bioinformatics/btac592.
[4] Liebig T, Vialard V, Opitz M. Connecting the Dots in Million-Nodes Knowledge Graphs with SemSpect. ISWC Posters Demos Ind. Tracks, 2017, p. Vol-1963:paper587.
[5] Henkel R, Wolkenhauer O, Waltemath D. Combining computational models, semantic annotations and simulation experiments in a graph database. Database (Oxford). 2015;2015. http://dx.doi.org/10.1093/database/bau130.
[6] Pall AS, Khaira JS. A comparative review of extraction, transformation and loading tools. Dbjournal.ro. http://dbjournal.ro/archive/12/12_5.pdf.
[7] Cheng KY, Pazmino S, Schreiweis B. ETL processes for integrating healthcare data - tools and architecture patterns. Stud Health Technol Inform. 2022;299:151–6.

[4] Neo4j: https://neo4j.com

[5] Neo4j Bloom: https://neo4j.com/docs/bloom-user-guide/current/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230256

Towards ETL Processes to OMOP CDM Using Metadata and Modularization

Yuan PENG[a,1], Elisa HENKE[a], Martin SEDLMAYR[a] and Franziska BATHELT[a]
[a] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*

Abstract. OMOP common data model (CDM) is designed for analyzing large clinical data and building cohorts for medical research, which requires Extract-Transform-Load processes (ETL) of local heterogeneous medical data. We present a concept for developing and evaluating a modularized metadata-driven ETL process, which can transform data into OMOP CDM regardless of 1) the source data format, 2) its versions and 3) context of use.

Keywords. Metadata, ETL, OMOP CDM, data harmonization, interoperability

1. Introduction

Data from clinics must be harmonized and standardized for unified data analysis in medical research. For this purpose, the Observational Medical Outcomes Partnership (OMOP) Common Data Models (CDM) [1, pp. 31–52] from Observational Health Data Sciences and Informatics (OHDSI) [1, pp. 1–9] community can be used. The development of required Extract-Transform-Load (ETL) processes [1, pp. 75–97] is time-consuming due to heterogeneous data formats and the lack of reusable ETL process components. In this work, we present an implementation concept of a generic ETL process for data harmonization in OMOP CDM, which is independent from 1) source data formats, 2) data specification versions, and 3) context of use.

2. Methods

Based on our monolithic ETL process from FHIR to OMOP [2] we developed a two-fold implementation using metadata and modularization. Metadata enables independence of ETL process from data formats and specifications by separating the mapping logic from it. Modularization supports reusability of ETL process components in different contexts. Transferability of the concept is evaluated in different medical use cases.

[1] Corresponding Author: Yuan Peng, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany; E-mail: yuan.peng@tu-dresden.de.

3. Results

Figure 1 shows the overall implementation concept of a modularized Metadata-driven (MDD) ETL process, which includes the encapsulation of the entire ETL process into nearly independent and reusable components.

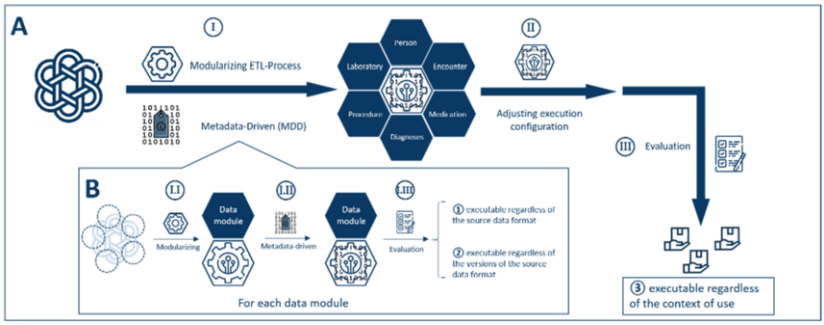


Figure 1. The concept of implementing a MDD-ETL-process (own illustration).

The evaluation concept consists of three parts (see Table 1). The evaluation criteria ① and ② are applied to each sub-modularized-MDD-ETL process (Figure 1, Part B). The evaluation criteria ③ is applied in Step III.

Table 1. Evaluation concept.

Nr.	Evaluation criterion	Considered success
①	ETL process is executable for different data formats	The implemented modularized MDD-ETL process is executed for German claims data format [3] and MI-I CDS FHIR format [4].
②	ETL process is executable for different versions.	The implemented modularized MDD-ETL process is executed for two versions of MI-I CDS FHIR format (e.g. v1.0 and v2.0).
③	ETL process is executable regardless of context of use	The implemented modularized MDD-ETL process is executed in at least two medical use cases.

4. Discussion and Conclusion

This work shows an initial developing concept of a modularized ETL process using metadata-driven approach to transfer different data formats to OMOP CDM, which includes an extensive evaluation design. Our next steps are 1) modularizing the current ETL-process, 2) choosing a metadata-driven approach through a scoping review for further implementation and 3) evaluating this concept on an international data set, e.g. US core data set.

References

[1] Observational Health Data Sciences and Informatics, *The book of OHDSI: Observational Health Data Sciences and Informatics*. San Bernardino, CA: OHDSI, 2019.

[2] Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. International Journal of Medical Informatics. 2023 Jan 1;169:104925.

[3] "Datensatzbeschreibung, InEK GmbH." https://www.g-drg.de/datenlieferung-gem.-21-khentgg/datenlieferung-gem.-21-abs.1-khentgg/dokumente-zur-datenlieferung/datensatzbeschreibung

[4] "The Medical Informatics Initiative's core data set | Medical Informatics Initiative." https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230257

Evaluation of Modeling Approaches for a Clinical Data Warehouse in a Highly Dynamic Environment

Stephan LORENZ[a,1], Richard GEBLER[a], Franziska BATHELT[a], Martin SEDLMAYR[a] and Ines REINECKE[a]
[a]*Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*

Abstract. The availability of clinical data for researchers is crucial for an improvement of healthcare and research. For this purpose, the integration, harmonization and standardization of healthcare-data from various sources in a clinical data warehouse (CDWH) is highly relevant. Our evaluation taking into account the general conditions and requirements of the project, led us to choose the Data Vault approach for the development of a clinical data warehouse at the University Hospital Dresden (UHD).

Keywords. Data Warehouse (DWH), Data Vault (DV), agile development

1. Introduction

A clinical data warehouse (CDWH) can be used to integrate and to harmonize medical data from various clinical sources in a central repository [1].The major requirements given by the Medical Informatics Initiative (MII) [2] and the general conditions of the implementing hospitals (i.e. available staff and budget) lead to the following question, addressed in this paper:

Which DWH modeling approach offers agile and flexible development of a CDWH according to volatile requirements of the MII for a small development team with limited budget, limited time and the aim to develop a framework for performant analytics of medical data?

2. Methods

We investigated the most suitable modeling approach for the CDWH based on the specific conditions at the University Hospital Dresden (UHD) and the requirements of the MII [2], conducting a selective literature research while considering the determining

[1] Corresponding Author: Stephan Lorenz, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany; E-mail: stephan.lorenz@mailbox.tu-dresden.de.

factors of the healthcare environment and the characteristics of heterogeneous medical data.

3. Results

A *flexible and agile development process* (Indicators for evaluation are marked *cursively*) is essential in order to make required adjustments caused by volatile requirements [2] and changing data sources without being limited by the necessity to redesign the entire data model.

According to Yessad and Labiod (2016) and Naame and Jovanovic (2016), the DV approach is the most suitable choice in terms of *philosophy*, *methodology* and *architecture*, *data integration* and *extraction, transform, and load (ETL)*, as well as *data modeling*. This approach is superior to the approaches of Inmon and Kimball in various aspects and can effectively address the challenges of *agile data-integration*, *-management*, and *-governance*. DV also follows an agile methodology, which results in faster *deployment times* and lower *costs of development*, compared to the other approaches, benefitting by its *flexibility* in handling *new analytical requirements* and *changes to the source data model(s)* [3,4].

4. Discussion and Conclusion

While the Data Vault approach offers many benefits such as flexibility, productivity, and scalability, it also has a higher level of complexity due to the high data standardization. In order to facilitate analysis and reporting, (dimensional) data marts (i.e. i2b2, OMOP CDM, MII core dataset, TriNetX) are used with the Data Vault approach in the CDWH of the UHD. These data marts need to adapt and reflect any changes in data sources through modification of the data delivery to the marts. It is important to consider this additional layer of complexity when implementing a CDWH using the DV approach. In order to identify and address potential bottlenecks in analytical performance, we regularly need to monitor and evaluate the performance and complexity of the CDWH.

References

[1] Hamoud A, Hashim AS, Awadh WA. Clinical data warehouse: a review. Iraqi Journal for Computers and Informatics. 2018;44(2).

[2] The Medical Informatics Initiative's core data set | Medical Informatics Initiative, (n.d.). https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set (accessed January 3, 2023).

[3] Yessad L, Labiod A. Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In2016 International Conference on System Reliability and Science (ICSRS) 2016 Nov 15 (pp. 95-99). IEEE. doi:10.1109/ICSRS.2016.7815845.

[4] Naamane Z, Jovanovic V. Effectiveness of Data Vault compared to Dimensional Data Marts on Overall Performance of a Data Warehouse System. International Journal of Computer Science Issues (IJCSI). 2016 Jul 1;13(4):16. doi:10.20943/01201604.1631.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230258

Modeling Patient Treatment Trajectories Using Markov Chains for Cost Analysis

Markus HAUG[a,1], Raivo KOLDE[a], Marek OJA[a] and Maarja PAJUSALU[a]
[a]*Institute of Computer Science, University of Tartu, Estonia*
ORCiD ID: Markus Haug https://orcid.org/0000-0003-0935-3307, Raivo Kolde https://orcid.org/0000-0003-2886-6298, Marek Oja https://orcid.org/0000-0002-3650-8194, Maarja Pajusalu https://orcid.org/0000-0001-9715-0678

Abstract. Electronically stored medical records offer a rich source of data for investigating treatment trajectories and identifying best practices in healthcare. These trajectories, which consist of medical interventions, give us a foundation to evaluate the economics of treatment patterns and model the treatment paths. The aim of this work is to introduce a technical solution for the aforementioned tasks. The developed tools use the open source Observational Health Data Sciences and Informatics Observational Medical Outcomes Partnership Common Data Model to construct treatment trajectories and implement these to compose Markov models for composing financial analysis between standard of care and alternatives.

Keywords. medical data, treatment trajectories, Markov chain, cost-effectiveness, OMOP CDM, heart failure

1. Introduction

The Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM) has been widely adopted for storing and analyzing medical observational data. It is currently being used in over 20 countries around the world. While the OHDSI CDM has a number of analytical tools available, there is a lack of tools specifically geared towards healthcare analytics with regards to costs and quality of life [1]. In response to this gap, we have developed easily distributable, transparent, and reproducibility-friendly research tools to fill this need.

2. Methods

To conduct healthcare analytics studies focused on costs and quality of life, we developed two R-packages. The first package, *Cohort2Trajectory*[2], utilizes the OHDSI CDM structure to generate patient trajectories by defining a target cohort and relevant

[1] Corresponding Author: Markus Haug, Research group of health informatics, https://health-informatics.cs.ut.ee; E-mail: markus.haug@ut.ee.
[2] https://github.com/HealthInformaticsUT/Cohort2Trajectory

state cohorts for the study. The tool resolves potential conflicts related to state overlap for each patient, offering various customization options for the decision-making process.

The second package, *TrajectoryMarkovAnalysis* [3], employs Markov modeling techniques on the resulting trajectory data to analyze the transition probabilities between health states. The package also retrieves and summarizes the costs associated with each trajectory, dividing them into state-specific costs. These Markov models and cost data are then used for both descriptive and cost-effectiveness analyses. The configurations for each study can be saved for reproducibility purposes.

3. Results and Discussion

To demonstrate the validity of our approach, we reproduced the study conducted by Thokala et al. [2], where they employed Markov models to calculate the incremental cost-effectiveness ratio for telemonitoring in heart failure patients. The results obtained using our developed tools on Estonian Health Insurance Fund data were similar to the original study. Specifically, the transition probabilities and monetary outputs of the Markov chains were comparable, leading to the same non-supporting conclusion about the feasibility of using telemonitoring for heart failure patients due to the high incremental cost. The calculated cost per quality-adjusted life year (QALY) exceeded the cost-effectiveness threshold agreed upon by healthcare institutions, making telemonitoring not a viable option for heart failure patients. Additionally, we employed survival analysis to compare the observed and generated trajectories, validating the learned Markov models.

The methods used to develop our tools have demonstrated their effectiveness and utility. By distributing our packages among OHDSI data partners, we can facilitate the reproducibility and transparency of any study defined using these tools. One potential challenge in conducting such studies is the availability and reliability of monetary data among data partners. Moving forward, we plan to focus on improving validation and comparison methods, including addressing potential issues related to the use of monetary data. Additionally, we aim to explore other methods for comparing patient treatment trajectories, including clustering trajectories with dynamic time warping and predicting future trajectories with neural networks.

The ability to analyze observational trajectories has numerous potential applications, including evaluating drug adherence and monitoring compliance or changes in the standard of care. In conclusion, our research emphasizes the importance of distributable and reproducible tools in healthcare for advancing the field.

References

[1] Petitti Kent S, Burn E, Dawoud D et al. Common Data Model Solutions: Evidence Generation for Health Technology Assessment. PharmacoEconomics. 2021; 39:275–285, doi: https://doi.org/10.1007/s40273-020-00981-9

[2] Thokala P, Dodd P, Baalbaki H, Brennan A, Dixon S, Lowrie K. Developing Markov Models From Real-World Data: A Case Study of Heart Failure Modeling Using Administrative Data. Value in Health. 2020; 23(6):743–750, doi: https://doi.org/10.1016/j.jval.2020.02.012

[3] https://github.com/HealthInformaticsUT/TrajectoryMarkovAnalysis

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230259

Representing Sex and Gender Information in Biomedical Research

Dagmar WALTEMATH[1a], Hilke BEELICH[a], Philipp TÖPFER[b], Corinna BERGELT[b], Agnes FLÖEL[c], Elpiniki KATSARI[d], Elke KRÜGER[e], Stine LUTZE[f], Michael SCHÖNER[g], Sebastian PASCHEN[h], Moritz ROLOFF[h], Hanne ROTH[g], Sylvia STRACKE[i]

[a] *Medical Informatics Laboratory, University Medicine Greifswald, Germany*
[b] *Medical Psychology, University Medicine Greifswald, Germany*
[c] *Neurology, University Medicine Greifswald, Germany*
[d] *Trauma Surgery, University Medicine Greifswald, Germany*
[e] *Medical Biochemistry, University Medicine Greifswald, Germany*
[f] *Dermatology, University Medicine Greifswald, Germany*
[g] *Graduate Academy, University Greifswald, Germany*
[h] *German Medical Students' Association, "Sex and Gender in Medicine"*
[i] *Internal Medicine, University Medicine Greifswald, Germany*

ORCiD ID: D Waltemath https://orcid.org/0000-0002-5886-5563; P Töpfer 0000-0002-0041-6538; C Bergelt 0000-0003-1413-1872; A Flöel 0000-0002-1475-5872; E Katsari 0000-0003-1956-8192; M Schöner 0000-0001-7538-7367; S Stracke 0000-0001-6200-4339

Abstract. In medicine and biomedical research, sex- and gender-related aspects are ubiquitous. If not considered adequately, a lower quality of research data can be expected together with a lower generalizability of study results with real-world settings. From a translational perspective, a lack of sex- and gender-sensitivity in acquired data can have negative implications for diagnosis, treatment (outcome and side effects), and risk prediction. To establish improved recognition and reward settings we set out to develop a pilot of systemic sex and gender awareness in a German medical faculty, with actions such as implementing equality in routine clinical practice and research, as well as in scientific practice (incl. science education). We believe that the change of culture will have a positive effect on research outcomes, lead to a rethinking in the scientific domain, foster sex- and gender-related clinical studies, and influence the design of good scientific practices.

Keywords. sex and gender, data life cycle, semantic enrichment

1. Introduction

In medicine, sex and gender-related aspects are ubiquitous and can be found in epidemiology, pathophysiology, clinical manifestations, psychological effects, disease progression, and response to treatment [1]. In this context, sex refers to biological constructs, whereas gender refers to social constructs. Sex and gender are genetic, biological, and environmental modifiers of acute and chronic diseases. Inadequate or insufficient consideration of sex and gender in medical research leads to reduced data quality and non-reproducibility of scientific outcomes [2]. From a translational

[1] Corresponding Author: Dagmar Waltemath, dagmar.waltemath@uni-greifswald.de.

perspective, a lack of sex- and gender-sensitivity in the data can have serious implications for diagnosis, treatment (outcome and side effects), and risk prediction [1].

2. InkE – Excellence through Inclusiveness in Medicine

InkE is a project at the University Medicine Greifswald, Germany, in close collaboration with Greifswald University's Faculty of Arts and Humanities, Graduate Academy and central gender equality office. The project team develops concepts, actions, and tools that all target the gender gaps in medical research and education.

One project goal is to change the system of academic recognition and reward. The current one-sided emphasis on traditional, quantifiable output indicators (e.g., number of publications, h-index and journal impact factor) is one of the causes of (implicit) gender bias. In order to achieve excellence, we aim for a renewed system of recognition and rewards that: (1) enables the diversification and vitalisation of career paths; (2) acknowledges the independence and individual qualities and ambitions of academics as well as recognising team performances; (3) emphasises quality of work over quantitative results; (4) encourages all aspects of open science; and (5) encourages high-quality academic leadership (see also https://recognitionrewards.nl).

A second goal is to encode sex and gender information in clinical and epidemiological studies. Good scientific practices for clinical and epidemiological studies exist and efforts have been made to harmonize and publish interoperable data models and metadata, e.g. [3]. However, a first analysis of data dictionaries for a sample of German studies revealed that only little information on participants' gender is recorded as of now. To complicate matters, in German, the term 'Geschlecht' simultaneously refers to both sex and gender – leading to unawareness or even lack of distinction between the two concepts. To overcome this situation quickly, we propose a fine-grained encoding of such information using domain ontologies [4]. Gender variables should be semantically enriched throughout the data life cycle, during data retrieval and exploration. Such provenance information will allow for gender-studies involving changes of gender in patients and consequences thereof in the context of medical care.

3. Acknowledgements

This work has been funded by BMBF, InkE Funding number: 01FP22G44A.

References

[1] Mauvais-Jarvis F, Merz NB, Barnes PJ, Brinton RD, Carrero JJ, DeMeo DL, De Vries GJ, Epperson CN, Govindan R, Klein SL, Lonardo A. Sex and gender: modifiers of health, disease, and medicine. The Lancet. 2020 Aug 22;396(10250):565-82.

[2] Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves science and engineering. Nature. 2019 Nov 7;575(7781):137-46.

[3] CO, Darms J, Shutsko A, Löbe M, Nagrani R, Seifert B, Lindstädt B, Golebiewski M, Koleva S, Bender T, Bauer CR. Facilitating Study and Item Level Browsing for Clinical and Epidemiological COVID-19 Studies. Studies in Health Technology and Informatics. 2021 May 1;281:794-8.

[4] Kronk CA, Dexheimer JW. Development of the Gender, Sex, and Sexual Orientation ontology: Evaluation and workflow. Journal of the American Medical Informatics Association. 2020 Jul;27(7):1110-5.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230260

A Categorial Structure for Identifying Physiological Measurement Observables

Anders THURIN[a,1]
[a] *Sahlgrenska University Hospital, Gothenburg, Sweden*
ORCiD ID: Anders Thurin https://orcid.org/0000-0002-5044-6364

Abstract. Patient measurements to characterize pathophysiological phenomena are often difficult to describe. A controlled terminology for this domain is needed, to describe methods, appropriate reference values and to report results. This is a proposal for a categorial structure for such a terminology

Keywords. Clinical Physiology, Observables, Terminology, Ontology.

1. Introduction

Functional measurements are important in management of many important diseases, and are often based on imaging and signal processing with secondary measurements and levels of postprocessing, resulting in complex nomenclature with risk of confusion. An example is LAVI(2dMOD), i.e. “left atrial volume estimated by a 2-dimensional method of disks model, indexed by body surface area”. There is a need for more systematic nomenclature of this domain. Some of this has been attempted in Clinical LOINC [1]. However, the structure of LOINC is adapted for clinical laboratory science and not optimal for functional patient measurements.

This paper outlines a proposed structure to adapt terminological systems better to identify physiological measurements, to facilitate unambiguous description of methods, diagnostic reporting and selection of reference values for interpretation of measurements.

2. Dedicated Kinds of Property

In laboratory medicine useful nomenclature is based on standardized description, based on controlled combinations of selected concepts (language independent) from prescribed categories, a so called **categorial structure**, as described in European standard EN 1614 [2]. Laboratory measurements are described as a **kind of property** of a **component** in a **system**, determined by a **method**. To describe physiologic measurements, however, a slightly more complex model is needed.

The essence of the proposed model is a small set of well understood entities, each with a small number of relevant measurable kinds of properties. Combining this with methods for measurement and challenges (condition during measurement), temporal relations and postprocessing operations (e.g. correction for body size) gives an economic

[1] Corresponding Author: Anders Thurin, E-mail: anders.thurin@vgregion.se

way to represent very complex observables. The classes comprising the model are here listed with some examples, but need to be further elaborated and defined, possibly in international standardization.

The proposed model is centered on a few types of structural/functional **entity**, each with a set of potential **functions** or **properties.** Some important examples are given in table 1.

Table 1. Entities with corresponding functional properties

Entity type	Kinds of properties (examples)	Comments, examples
Wall	thickness, motility, strain, strain rate, echogenicity, radiolucency and its heterogeneity, continuity	Muscular wall of heart, urinary bladder
Lumen (inside a hollow structure)	diameter, area (in a sectional plane), volume, stroke volume, volume fraction, fractional shortening, presence of pathological structure, presence of flow	Heart ventricle, inside of artery, airways
Tubular structure	diameter, area (cross section), plaque (presence, size, echolucency, heterogenity, shape), patency, tortuosity	Blood vessel, ureter,
Valve	thickness, mobility, patency, regurgitation, prosthesis	Heart valves, venous valves
Solid structure	presence, size, diameter, volume, echogenicity, radiolucency, uptake of tracer/contrast, rhythmicity (e.g. heart rate), mobility	Kidney, thrombus,
Patient	height, weight, body surface area, exercise capacity	Properties applicable to patient as a whole
Flow	presence (patency), direction, velocity, averaged velocity integral, volume flow, pulsatility, wave reflection, acceleration, deceleration	Arterial, venous, respiratory, flow across heart valve

Several other dimensions are useful for a systematic description:

Method used for measurement/data collection is of obvious importance, and includes technical equipment used, such as ultrasound, magnetic resonance, scintigraphy, and sometimes also **view/orientation** (such as parasternal view of heart, anteroposterior view of pelvis).

Timing Many properties have a relation to a time axis, e.g. velocity, acceleration, or relation to events such as heart beat (systole, diastole).

Challenge – some observables can only be described by their relation to a challenge, such as body position, muscular contraction, breathing, external compression, pharmaceutical challenge.

Post processing – frequently observables are adjusted by calculation of indexes, dividing by body mass, or other measurements, and sometimes advanced processing by imaging algorithms, and this often is an important part of the description.

References

[1] Vreeman DJ, McDonald CJ, Huff SM. LOINC(R) - A Universal Catalog of Individual Clinical Observations and Uniform Representation of Enumerated Collections. Int J Funct Inform Personal Med. 2010;3(4):273-91.

[2] CEN. EN 1614:2006 Health informatics - Representation of dedicated kinds of property in laboratory medicine. 2006.

Section 7

Natural Language Processing

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230262

Automated ICF Coding of Rehabilitation Notes for Low-Resource Languages via Continual Training of Language Models

Kevin ROITERO[a], Andrea MARTINUZZI[b], Maria Teresa ARMELLIN[b], Gabriella PAPARELLA[b], Alberto MANIERO[b] and Vincenzo DELLA MEA[a,1]
[a] *Dept. of Mathematics, Computer Science and Physics, University of Udine, Italy*
[b] *Dept. of Conegliano, IRCCS "E. Medea" Scientific Institute, Conegliano, Italy*
ORCiD ID: Kevin Roitero https://orcid.org/0000-0002-9191-3280,
Andrea Martinuzzi https://orcid.org/0000-0002-0319-3579,
Vincenzo Della Mea https://orcid.org/0000-0002-0144-3802

Abstract. The coding of medical documents and in particular of rehabilitation notes using the International Classification of Functioning, Disability and Health (ICF) is a difficult task showing low agreement among experts. Such difficulty is mainly caused by the specific terminology that needs to be used for the task. In this paper, we address the task developing a model based on a large language model, BERT. By leveraging continual training of such a model using ICF textual descriptions, we are able to effectively encode rehabilitation notes expressed in Italian, an under-resourced language.

Keywords. Language Models, ICF, Rehabilitation, Continual Training

1. Introduction

The International Classification of Functioning, Disability and Health (ICF) is a classification of functioning conditions developed by the World Health Organization (WHO) [1]. The concept of functioning on which ICF is built upon is that of a "dynamic interaction between a person's health condition, environmental factors and personal factors". This overall model has been translated into a classification covering all the main components of functioning, namely Body Structures and Functions, Activities & Participation, and Environmental Factors.

Coding natural language (NL) notes with ICF is recognized as difficult task [2] and also with low inter-observer agreement [3]. Yet, the need for coding electronic health records also with ICF has been recognized as useful [4]. Thus, the use of tools to support this task is welcomed, although not yet researched as much as for other biomedical classifications like ICD, e.g., in [5,6,7,8].

Training models for ICF coding presents a main difficulty if compared with other biomedical classifications. In fact, while other terminologies and classifications include specialized terms and concepts (disease and procedure names, anatomy parts, etc.), ICF, in particular in its Activities & Participation and Environmental Factors components,

[1] Corresponding Author: Vincenzo Della Mea, E-mail: vincenzo.dellamea@uniud.it

give specific meaning within the ICF biopsychosocial framework to commonly used terms and concepts (e.g., Walking, Washing oneself, Food, Temperature, etc.). However, there have been seminal attempts to automated ICF coding ([9]), and recently attention resumed ([10,11]), also and notably in an under-resourced language like Dutch ([12]).

In the present paper, we propose a methodology for automated coding of NL texts using ICF, in an under-resourced language as Italian, thanks to the availability of a real-world dataset provided by a rehabilitation clinic in North-Eastern Italy. A preliminary pilot experiment probed the possibility of automatically code NL strings to the correct ICF code in 2 chapters of ICF: one from the Body Functions domain (functions related to movement) and one from the Activity & Participation domain (mobility).

2. Methods

2.1. Dataset

Recognizing the power of ICF in univocally describing the various aspects of human functioning, the use of codes from the whole ICF coupled with the appropriate qualifiers was systematically introduced from 2014 in a tertiary neurorehabilitation centre in Northern Italy. To ease initial use and to allow more accurate description of the content of the ICF entity used, each code is associated with a natural language free text where the professional describes in lay terms what he/she observes.

Since 2014, over 2000 complete functioning profiles of patients admitted to the centre have been stored and are now available for analysis. Health conditions span from cerebral palsy to Parkinson and patients age from 0 to > 80 years. These projects contain detailed descriptions attached to selected ICF codes of any level and the attached qualifiers. Retrievable data include: ICF codes and qualifiers used; Natural language description of the item; Linkage between ICF categories and evaluation tools.

Additional data for continual training was collected with the aim of focusing on ICF expressions. Since the kind of notes found in the rehabilitation dataset are not available from other sources, we identified some relatively suitable, yet sub-optimal, Italian texts in the ICF classification itself [1] and the Wikipedia article describing it.

The dataset was pre-processed as follows. We started with the full dataset and we removed the classes which appeared only once, which are 129; this left us with 473 classes that have associated at least two instances. Then, we subset the dataset in two different ways, as follows. To create the former dataset, we selected all the instances that belong either to the B7 or D4 chapters (associated to 107 classes). These classes are related to the important area of mobility from two different points of view: Body Functions (B7) and Activities & Participation (D4). Then, we split the dataset into the training and test sets performing stratified sampling; this left us with 6,650 instances in training and 1,751 in test. To create the latter dataset, we augmented the training set of the first one with instances not belonging to the B7 or D4 classes; the rationale behind this choice is that by leveraging more data the models we use can better tell apart the B7 and D4 classes, both each other and also from all the other classes. This second dataset share the test set with the former dataset and has 30, 438 instances in training.

2.2. Model training and inference

To develop and train our models we rely on both the PyTorch and HuggingFace frameworks. The models were trained and tested on a Linux server equipped with Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz, 64GB of RAM, and 2x Nvidia Geforce RTX 3090 GPUs. The trained models are available for research purposes.

BERT [13] is a transformer based model pre-trained on large corpora of multilingual text using a self-supervised training procedure. More in detail, the model has been trained on the textual corpora without the usage of any human annotation or feedback thus by relying on a sampling technique which is used to generate inputs and labels from the raw text. The model is trained with two objectives, Masked Language Modeling (MLM) and Next Sentence Prediction (NSP): the former receives in input a sentence and randomly masks a percentage (usually 15%) of the input tokens and asks the model to predict the masked tokens, while the latter takes two sentences in input which are either one after the other in the original text or not with a given probability (usually 50%) and asks the model to predict such relation between sentences (i.e., whether they were following each other in the input text or not). Note that such two training procedures are different from the classical ones used by recurrent networks which receive in input the words or tokens one after the other, and also from auto-regressive models like T5 [14], which receives mask only future tokens. The generated pre-trained model can be then fine-tuned on a variety of downstream tasks such as classification, regression, etc.

In this work, we rely on the *bert-base-multilingual-cased* model, a 110 million parameters model pre-trained on the largest Wikipedia dump in 104 different languages. Details on the training procedure are as follows: for the MLM training procedure 15% of the tokens have been masked; among those, in 80% of the cases the masked tokens are replaced with "[MASK]", in 10% of the cases thy are replaced with a random token, and in remaining 10% of the cases the input sentence is not altered. For NSP the probability for sampling subsequent sequences is set to 50%.

Starting from the base model, we developed two model variations; in the former we simply attach a classification head to the model to fine-tune it to our dataset, while for the latter we perform "continual training" before attaching the classification head: we continue the pre-training procedure of the model on ICF text for additional 100 epochs for the MLM objective keeping the sampling parameters as for the original model. We monitored the model losses to make sure the training procedure was stable and we did not encounter over-fitting or damage model weights. We did not use the NSP training procedure for the final model because we found it does not provide any increase in the effectiveness of the model; this is consistent with literature results, e.g., the training procedures implemented in the HuggingFace framework for the BERT based models.

3. Results

Table 1 shows the effectiveness of the proposed approach. The first three columns of the table detail the base model used (i.e., BERT), the training objective (i.e., either the usage of the plain pre-trained model or the model where we performed continual training), and the training dataset (i.e., the 6,650 instances with B7-D4 or the full dataset consisting of 30,438 instances), while the latter columns detail the values for the Accuracy and F1 scores, computed at different cutoffs; note that macro-average weights all classes equally independently of their frequency, as it is obtained by computing the metric independently

for each class and then taking the average, while micro average aggregated the contributions from the different classes to compute the metric, hence it is usually preferred in the multi-class scenario.

Table 1. Effectiveness of the proposed approach.

Base model	Training objective	Training dataset	Accuracy at 1	Accuracy at 3	Accuracy at 5	F1 micro	F1 macro	F1 weighted
BERT	pre-trained	B7-D4	.683	.864	.905	.683	.179	.607
BERT	continual	B7-D4	.709	.882	.925	.709	.226	.654
BERT	pre-trained	Full	.760	.910	.943	.760	.231	.718
BERT	continual	Full	.779	.917	.949	.779	.260	.746

As we can see from the table, both the training objective and the training dataset have an impact on the effectiveness scores. More in detail, as we can see by comparing respectively the first and second, and the third and fourth row of the table, the continual training objective increases the effectiveness scores; this is probably due because the further pre-training of the model allows to capture the context and semantics of ICF.

Furthermore, as we can see by comparing the first and last two pair or rows in the table, the training set employed has an effect of the effectiveness scores; in particular, the full dataset achieves higher effectiveness scores. This is probably due because the additional training instances allow the model to better tell apart the test classes, both one another as well as with all the other classes.

Overall, we can see that the model which achieves higher effectiveness scores is the one obtained by leveraging both continual training as well as the full training dataset.

4. Discussion

The results of this preliminary experiments seem to suggest the feasibility of automated ICF coding from text notes, at least for the specific subset of codes related to mobility. In the best case, Accuracy@1 is 0.779, which could be the performance of a totally automated system. A coding support system giving the coder three options among which to choose could attain a higher accuracy (0.917).

However, when looking at macro-averaged F1, the results suggest caution. The substantially lower values are given by the fact that ICF include a fair number of codes, many of which rarely used. So, performance is high on frequent codes only. Similar works in the past obtained higher F1 scores, but focusing on a smaller set of target codes. For example, in [11] two datasets with 13 and 16 ICF codes respectively were used; in [9] coding involved five codes; in [12] 4 main ICF categories were mentioned. In our dataset we removed codes used only once, but still the total number of used codes was 107, and this, with the selected models, decreased performance.

One limitation of the present work is the fact that the dataset, coming from a clinical setup, includes codes each provided by a single coder, thus with no measure of inter-observer agreement among human coders and, possibly, coding mistakes.

5. Conclusions

The preliminary experiment on ICF coding provides some feasibility evidence, with still some room for improvement for less represented codes.

Future works to overcome this limitation include:

- experimentation of zero- or few-shots learning methods to better cover the least used codes;
- identification of possible sources of text for continual training, which do not need to be coded in ICF, but still mentioning relevant concepts;
- if datasets in other languages become available, experimentation with large transformer-based multilingual models with multiple training languages and/or high generalization (i.e., zero- and few-shot learning) capabilities.

Acknowledgements

We thank for the support AM received from the Italian Ministry of Health (RC2022-23).

References

[1] World Health Organization. International Classification of Functioning, Disability and Health. WHO; 2001.
[2] Schuntermann MF. The implementation of the International Classification of Functioning, Disability and Health in Germany: experiences and problems. International Journal of Rehabilitation Research. 2005;28(2):93-102.
[3] Okochi J, Utsunomiya S, Takahashi T. Health measurement using the ICF: test-retest reliability study of ICF codes and qualifiers in geriatric care. Health and quality of life outcomes. 2005;3(1):1-13.
[4] Ustun TB, Chatterji S, Kostansjek N, Bickenbach J. WHO's ICF and functional status information in health records. Health Care Financ Rev. 2003;24(3):77-88.
[5] Gobeill J, Ruch P. Instance-based Learning for ICD10 Categorization. In: CEUR Workshop Proceedings 2125, CEUR-WS.org; 2018.
[6] Li F, Jin Y, Liu W, Rawat BPS, Cai P, Yu H. Fine-Tuning Bidirectional Encoder Representations From Transformers (BERT)-Based Models on Large-Scale Electronic Health Record Notes: An Empirical Study. JMIR Med Inform. 2019 Sep;7(3):e14830.
[7] Roitero K, Portelli B, Popescu MH, Della Mea V. DiLBERT: Cheap Embeddings for Disease Related Medical NLP. IEEE Access. 2021;9:159714-23.
[8] Falissard L, Morgand C, Ghosn W, Imbaud C, Bounebache K, Rey G. Neural Translation and Automated Recognition of ICD-10 Medical Entities From Natural Language: Model Development and Performance Assessment. JMIR Med Inform. 2022 Apr;10(4):e26353.
[9] Kukafka R, Bales ME, Burkhardt A, Friedman C. Human and automated coding of rehabilitation discharge summaries according to the International Classification of Functioning, Disability, and Health. J Am Med Inform Assoc. 2006;13(5):508-15.
[10] Newman-Griffis D, Fosler-Lussier E. Automated Coding of Under-Studied Medical Concept Domains: Linking Physical Activity Reports to the International Classification of Functioning, Disability, and Health. Front Digit Health. 2021 Mar;3.
[11] Newman-Griffis D, Maldonado JC, Ho PS, Sacco M, Silva RJ, Porcino J, et al. Linking Free Text Documentation of Functioning and Disability to the ICF With Natural Language Processing. Front Rehabil Sci. 2021 Nov;2.
[12] Meskers CGM, van der Veen S, Kim J, Meskers CJW, Smit QTS, Verkijk S, et al. Automated recognition of functioning, activity and participation in COVID-19 from electronic patient records by natural language processing: a proof- of- concept. Ann Med. 2022 Dec;54(1):235-43.
[13] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.
[14] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res. 2020;21(140):1-67.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230263

Evaluating the Portability of Rheumatoid Arthritis Phenotyping Algorithms: A Case Study on French EHRs

Thibaut FABACHER[a,b,c,d,1], Erik-André SAULEAU[a,b], Noémie LECLERC DU SABLON[a], Hugo BERGIER[a], Jacques-Eric GOTTENBERG[a], Adrien COULET[c,d,2] and Aurélie NÉVÉOL[e,2]
[a] *University hospital of Strasbourg, France*
[b] *ICube Laboratory, Strasbourg, France*
[c] *Inria Paris, France*
[d] *Centre de Recherche des Cordeliers, Inserm, Université Paris Cité, France*
[e] *Université Paris-Saclay, CNRS, LISN, Orsay, France*

Abstract. Previous work has successfully used machine learning and natural language processing for the phenotyping of Rheumatoid Arthritis (RA) patients in hospitals within the United States and France. Our goal is to evaluate the adaptability of RA phenotyping algorithms to a new hospital, both at the patient and encounter levels. Two algorithms are adapted and evaluated with a newly developed RA gold standard corpus, including annotations at the encounter level. The adapted algorithms offer comparably good performance for patient-level phenotyping on the new corpus (F1 0.68 to 0.82), but lower performance for encounter-level (F1 0.54). Regarding adaptation feasibility and cost, the first algorithm incurred a heavier adaptation burden because it required manual feature engineering. However, it is less computationally intensive than the second, semi-supervised, algorithm.

Keywords. Natural Language Processing, Phenotyping, Rheumatoid Arthritis

1. Introduction

Electronic Health Records (EHRs) enable secondary use of hospital data, and in particular the design and conduct of clinical studies. A first step of such studies is the definition of a cohort of patients who share a specific condition. This task is referred to as electronic phenotyping and is often more complex than a simple query [1,2]. Searching a unique phenotypic trait in EHRs usually requires covering both structured fields and unstructured texts in a specific time frame. Furthermore, phenotyping algorithms may not transfer well from one clinical setting to another because of variations in data collection, clinical practice, coding of medical acts, policies, or language used in clinical notes. In general, cohort definitions rely on phenotyping at the patient level, but a finer granularity is necessary in some cases, especially when monitoring chronic diseases, since a same patient's encounters (i.e., visits) may or may not be related to the phenotype.

[1] Corresponding Author: Thibaut FABACHER, E-mail: thibaut.fabacher@chru-strasbourg.fr.
[2] Equal contribution.

In this work, we study the portability of phenotyping algorithms for Rheumatoid Arthritis (RA), a long-term autoimmune pathology that primarily affects joints. We explore with RA because it is relatively frequent, it poses clinical questions (e.g., what is a patient's prognosis, or best treatment options) and because phenotyping algorithms for RA have been described in the literature [3,4]. In particular, we compare three RA phenotyping algorithms on unseen EHR data to address the following questions: Which one is the most efficient in terms of performance and speed? Which one is easier to adapt to a new hospital, here the University Hospital of Strasbourg (UHS), France? Which one is prone to performance decrease when transferred?

2. Materials and Methods

Records of patients with encounters in 2015-2020 and high probability of RA are extracted from the UHS health information system. Specifically, we select patients with at least one ICD-10 code related to RA and one reference to RA in a clinical text over this time period. ICD-10 codes for RA are M060*, M068*, M069*, M058*, M059*, M053*, M050*; and detection of RA in French texts is performed with the regex pol(i|y)arth?rites? *?rh?umat, searching for "polyarthrite rhumatoïde" and its variations due to typos. Clinical notes (discharge summaries, progress reports, etc.), diagnostic codes (ICD-10), drug prescriptions and laboratory results were extracted for these patients. We excluded encounters only associated with ICD-10 codes or prescriptions. We excluded notes with content limited to ICD-10 codes or antecedent. This study is listed on UHS study register and follows the hospital clinical study protocol.

Data are split in three patient sets. 11% are randomly selected to form the exploration set, which is used to evaluate regex from Carroll's algorithm. The remaining 89% is split in a customized way so 85% constitutes our train set and 4% our test set. The customized sampling strategy is performed to obtain balanced groups of patients in the test set. Both train and exploration sets are used to train the PheVis Algorithm. The test set is annotated and used to evaluate all different methods. For the evaluation of phenotyping algorithms, we manually annotated our test set both at the patient and encounter levels. Accordingly, each encounter is annotated by two distinct individuals (among one rheumatologist and two public-health physicians). Each encounter is annotated with one of the following four labels: RA+ if the encounter is due to RA: diagnosis, assessment of disease progression, therapeutic management of the disease, management of complications of the disease; RA- if the encounter is not related to RA, even if the patient has an active RA; doubtful if the encounter cannot be confidently classified in relation to RA. To reach consensus, encounters annotated with two distinct labels are identified and discussed during a meeting. Doubtful encounters are ultimately labeled as RA- for method assessments, as the classification task that is evaluated is binary. For patient-level annotations, if a patient has at least one encounter labeled as RA+, she/he is labeled as RA+ at the patient-level, and as RA- otherwise. **Baseline Algorithm:** Encounters are classified RA+ if they have at least one ICD-10 code for RA and at least one mention of RA in a clinical text. Matching with ICD-10 codes is performed according to the list of ICD-10 codes described in the Data Collection paragraph. Matching with clinical texts is performed with a dictionary-based NER tool, named IAMsystem [4]. In addition, we performed two filtering to avoid false positives, referred to as contextualization in the following. First filtering excludes parts of the text concerning medical history. To this aim, we use a house-made algorithm, for section segmentation. Second filtering consists

in taking into account the context of RA mentions in clinical texts. To this aim, we use FastContext [5] and more specifically its implementation named IAMFastContext. With this tool, mentions of RA which are negated, hypothetical, historical or related to relatives or other persons are filtered out. **Carroll's Algorithm** [6,7] uses pretrained penalized logistic regression. Carroll et al. provide parameters of the regression, to enable the reuse of the classifier on new data. This algorithm takes as input structured data (ICD-9 codes, drug prescriptions, lab results) and named entities found in clinical texts with a set of regex. To adapt this algorithm to the UHS, the ICD-9 codes are manually converted to ICD-10, drug prescriptions and laboratory results of the UHS are adapted to be consistent with the classifier, and finally, the set of regex provided in English is adapted to French. The original regex, their translations and adaptation to French are available at https://gitlab.inria.fr/heka/ra_phenotyping/. We apply Carroll's algorithm with coefficients provided in the original article, and a probability threshold of 0.5 for classification of RA patients. **PheVis Algorithm** [4] leverages the PheNorm [8] method to classify patients following a semi-supervised approach. It classifies not only at the patient level, but also at the encounter level. PheVis is a two-stage approach, as it relies first on the definition of a silver standard of automatically annotated examples, that is used, second, to train a supervised model. PheVis takes as input ICD-10 codes and UMLS entities automatically extracted from EHR narratives with NLP. Our adaptation of the PheVis algorithm relies on the IAMSystem for entity extraction and normalization [9] to ensure comparability with the original PheVis study [4]. To test the portability of PheVis to the UHS setting, we test the best hyperparameters reported by PheVis authors.

Evaluation metrics Phenotyping is assessed with precision, negative predictive value (NPV), specificity, recall (or sensitivity), balanced accuracy, accuracy, F1 score and Area Under the ROC Curve (AUC). Confidence intervals are computed using bootstrap. Experiments use R version 4.1 and a personal computer under Windows 10, with 64Gb of memory and an Intel(R) Xeon(R) CPU E3-1245 v5.

3. Results

We found 4,100 patients with at least one ICD-10 code for RA and one reference to RA in narratives. The 410 patients with the most recent first encounter were selected as a validation set for future work. Remaining 3,690 patients were split in 410 (11%), 3,140 (85%) and 140 (4%) patients to constitute exploration, train and test sets. These include 3,826, 33,007 and 1,552 distinct encounters with at least one clinical text. Of the 1,552 encounters selected for manual annotation, after consensus on the annotation, 1,146 were classified as RA-, 358 as RA+ and 48 as doubtful. Inter-annotator agreement was substantial (Cohen's kappa = 0.80). At the patient level (n=140), 52 (37%) were classified as RA+ and 88 (63%) as RA-. Table 1 and 2 summarize the results of our evaluation of phenotyping algorithms. Baseline Algorithm For hospital encounters classification, F1 was 0.60 [0.55-0.64] and 0.61 [0.55-0.65] respectively for the basic algorithm without and with contextualization. For patient classification, F1 was 0.69 [0.60-0.78] and 0.73 [0.65-0.83] without and with contextualization of named entities. Contextualizing NERs improves performance, in particular precision. Carroll's Algorithm For patient classification, F1 is 0.82 [0.75-0.90]. Results are similar to those of Carroll et al. [7] (AUC=0.94 vs. AUC=0.95), with a higher specificity (0.82 vs. 0.65) and a lower precision (0.75 vs. 0.90). Carroll's algorithm is not available for encounter-level phenotyping. Phevis Algorithm For patient classification, results are lower than

those reported in Phevis paper (AUC=0.87 vs 0.94), F1 =0.68 [0.59-079]). F1 is lower, 0.54 [0.50-0.58] for encounter-level. The baseline algorithm is fairly easy to implement. ICD-10 codes are easy to extract from structured data. Regex matching is also fast, taking less than 20 minutes in our setting. Implementing the Carroll's algorithm took longer. About two working days was necessary to translate regex from English to French. It took one week to examine and modify regex with the exploration set. Searching to match all regex on the test set took about one hour. Implementing the logistic regression took half a day and the execution time of the logistic regression is almost instantaneous. The implementation of PheVis algorithm took more time. For data preparation, the NER with IAMsystem algorithm, took about two days to run on the exploration, train, and test datasets. Training a model took about 10 minutes. Once the classification algorithm is trained, application on new data is fast and takes about one minute.

Table 1. Performances for RA phenotyping at the patient level. PheVis setting is ω= 5, half-life = Inf

Methods	Prec.	NPV	Spe.	Rec.	bal Acc.	Acc.	F1	AUC
ICD-10 alone (≥1 code)	0.55	0.91	0.56	**0.90**	0.73	0.69	0.68 (0.58-0.77)	N/A
Baseline algo.	0.58	0.88	0.64	0.85	0.73	0.71	0.69 (0.60-0.78)	N/A
Baseline algo., plus context	0.67	0.87	0.76	0.81	0.77	0.78	0.73 (0.65-0.83)	N/A
Carroll's algo.	0.75	**0.94**	0.82	**0.90**	**0.84**	**0.85**	**0.82 (0.75-0.90)**	0.94 (0.89-0.99)
PheVis	0.62	0.84	0.72	0.77	0.73	0.74	0.68 (0.59-0.79)	0.87 (0.81-0.93)
Caroll, reported *et al.*	**0.90**	N/A	0.65	N/A	N/A	N/A	N/A	**0.95**
Phevis, reported *et al*	0.65	0.96	**0.94**	0.74	N/A	N/A	N/A	0.943

Table 2. Performances for RA phenotyping at the encounter level. PheVis setting is ω= 5, half-life = Inf

Methods	Prec.	NVP	Spe.	Rec.	bal Acc.	Acc.	F1	AUC
Baseline algo.	0.62	0.88	0.89	0.58	0.75	0.82	0.60 (0.55-0.64)	N/A
Baseline algo.	0.66	0.87	0.92	0.55	0.76	0.83	0.60 (0.55-0.64)	N/A
Baseline algo., plus context	**0.71**	0.87	**0.94**	0.53	**0.79**	**0.84**	**0.61 (0.56-0.65)**	N/A
PheVis	0.43	**0.89**	0.72	**0.71**	0.66	0.72	0.54 (0.50-0.58)	0.79 (0.76-0.82)

4. Discussion

Porting phenotyping algorithms from one setting to another remains a challenge. On UHS data, PheVis appears to have lower performance to Carroll's and baseline algorithms for patient phenotyping. Our adaptations of algorithms yield performance slightly lower than those reported in the literature (Table 1). The better results achieved so far may be due to the definition of RA+ patients, i.e. those with a history of RA or active RA. A recent study makes the same observation about the difficulty of adaptation and highlights the difficulty of defining the phenotyping task [10]. One originality of our study is the evaluation of algorithms at the encounter level. Although the authors of PheVis considered phenotyping encounters, their algorithm was evaluated only at the patient level. Our study suggests that PheVis is not superior to other algorithms at the encounter level. The rather good results we observed with the baseline algorithm, regarding what is reported in literature, may be attributed to an improvement of the coding in French hospitals. For patient phenotyping, the majority of false positive predictions are due to our definition of RA+ patients. Majority of the false positives are patients with a history of RA. Contextualization improves precision by removing part of the patient's history, but redundancies between encounters are found even if they are not directly related to the chronic disease [11]. Reducing this should reduce the number of

false positives. In this initial study of RA phenotyping in French EHRs, our goal was to use state-of-the-art algorithms validated in previous studies on new patient data. For Carroll's algorithm, differences between languages in terms of sentence construction make it difficult to translate complex regex from one language to another. PheVis phenotyping performance is good at the patient level. Adjusting the hyperparameters to the local context and pathology would allow for even better results. Unsurprisingly, encounter-level performance is underwhelming, as it is a harder task. Results may be improved by using more complex unsupervised machine learning classifiers like Phe2vec [12]. The Phevis method requires as little expert knowledge as the rule-based algorithm, this method can be used across hospitals, provided a suitable NER algorithm is available. Carroll's algorithm is more difficult to adapt to another phenotype. Feature extraction could be easily adapted; however, the availability of an annotated dataset to train the classifier is a bottleneck.

5. Conclusion

The two algorithms tested for RA phenotyping are transferable to the context of the UHS. In both cases, adaptation required a significant amount of time, whether for the translation of regular expressions or the implementation of a NER algorithm. The performance gain compared to a baseline algorithm relying solely on ICD-10 codes is surprisingly low. Previous studies did not always consider the baseline in their evaluation and encounter-level phenotyping needs to be better considered. More advanced machine learning algorithms, taking into account redundancy or more specific silver standard, could improve performance in future work.

References

[1] Newton KM, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. JAMIA 2013. doi:10.1136/amiajnl-2012-000896.
[2] Weng C, Shah NH, Hripcsak G. Deep phenotyping: Embracing complexity and temporality-Towards scalability, portability, and interoperability. J Biomed Inform 2020. doi:10.1016/j.jbi.2020.103433.
[3] Carroll RJ, Eyler AE, Denny JC. Intelligent use and clinical benefits of electronic health records in rheumatoid arthritis. Expert Rev Clin Immunol 2015. doi:10.1586/1744666X.2015.1009895.
[4] Ferté T, Cossin S, Schaeverbeke T, Barnetche T, Jouhet V, Hejblum BP. Automatic phenotyping of electronical health record: PheVis algorithm. J Biomed Inform 2021. doi:10.1016/j.jbi.2021.103746.
[5] Chapman WW, Hillert D, Velupillai S, Kvist M, Skeppstedt M, Chapman BE, et al. Extending the NegEx lexicon for multiple languages. Stud Health Technol Inform 2013;192:677–81.
[6] Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis Care Res 2010. doi:10.1002/acr.20184.
[7] Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. JAMIA 2012. doi:10.1136/amiajnl-2011-000583.
[8] Yu S, Ma Y, Gronsbell J, Cai T, Ananthakrishnan AN, Gainer VS, et al. Enabling phenotypic big data with PheNorm. J Am Med Inform Assoc JAMIA 2018;25:54–60. doi:10.1093/jamia/ocx111.
[9] Cossin S, Jouhet V, Mougin F, Diallo G, Thiessard F. IAM at CLEF eHealth 2018: Concept Annotation and Coding in French Death Certificates 2018. doi:10.48550/arXiv.1807.03674.
[10] Zheng HW, Ranganath VK, Perry LC, Chetrit DA, Criner KM, et al. Evaluation of an automated phenotyping algorithm for rheumatoid arthritis. J Biomed Inform 2022. doi:10.1016/j.jbi.2022.104214.
[11] Digan W, et al. Evaluating the Impact of Text Duplications on a Corpus of More than 600,000 Clinical Narratives in a French Hospital. MEDINFO 2019 2019:103–7. doi:10.3233/SHTI190192.
[12] De Freitas JK, Johnson KW, et al. Phe2vec: Automated disease phenotyping based on unsupervised embeddings from electronic health records. Patterns 2021. doi:10.1016/j.patter.2021.100337.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230264

Learning to Classify Medical Discharge Summaries According to ICD-9

Leonardo MOROS[a,1], Jérôme AZÉ[a], Sandra BRINGAY[a,b], Pascal PONCELET[a], Maximilien SERVAJEAN[a,b] and Caroline DUNOYER[c,d]
[a] *LIRMM UMR 5506, University of Montpellier, CNRS, Montpellier, France*
[b] *AMIS, Paul-Valéry University, Montpellier, France*
[c] *Medical Information Department, CHU Montpellier, Montpellier, France*
[d] *IDESP, UMR UA11, INSERM - University of Montpellier, Montpellier, France*

Abstract. Context: We present a post-hoc approach to improve the recall of ICD classification. **Method:** The proposed method can use any classifier as a backbone and aims to calibrate the number of codes returned per document. We test our approach on a new stratified split of the MIMIC-III dataset. **Results:** When returning 18 codes on average per document we obtain a recall that is 20% better than a classic classification approach.

Keywords. Supervised learning, constrained optimization, NLP

1. Introduction

Healthcare professionals meticulously record each patient's hospital visit via structured and semi-structured documents, which contain information about treatments, procedures and diagnoses. For financial reasons, healthcare institutions must associate billing codes from the International Classification of Diseases (ICD) with the aforementioned hospital visits. ICD coding is currently done manually by specialists. It is a very complex, tedious, subjective, costly, time-consuming and error-prone task. For example, in the United States, the cost of coding a single discharge summary is estimated to be about $172 [1]. The desire to limit costs has made the field of automatic coding a very active area of research. In recent years, the best results have been achieved by approaches using deep learning models. Mullenbach et al. [2] propose an approach combining convolutional neural networks and attention. Shi et al. [3] combine character level and sentence level LSTMs to code specific sections of the discharge summary. Teng et al. [4] combine convolutional neural networks, graph embeddings for code hierarchy, and attention. More recently, transformer based pre-trained language models have also been used. However, they do not outperform recurrent neural networks [5] in clinical coding. In this work, we propose a classifier that incorporates a budget approach. It is important to note that an error is the result of either an omitted label or of a wrong prediction. In the former case, the clinical coder must read the entire document to identify the omitted label. In the latter case, to invalidate the label, the coder only focuses on the parts of the document used by the model to make its prediction. In this context, our goal is to maximize recall.

[1] Corresponding Author: Leonardo Moros, E-mail: leonardo.moros@lirmm.fr.

The majority of papers evaluate their approaches on the existing MIMIC-III splits [2]. These splits have important stratification issues, which complicate the evaluation and comparison of methods. We evaluate the architecture of our classifier by comparing it to LAAT [6] on a new split of the MIMIC-III dataset that guarantees the stratification of labels in all the subsets.

2. Method

Let $\mathcal{X}$ be the input space (the discharge summaries associated with each patient) and $\mathcal{Y}$ the nodes of the ICD-9 hierarchy. The cartesian product $\mathcal{X} \times P(\mathcal{Y})$ is a probability space with a joint probability measure $P_{X,Y}$ where $Y \in \{0,1\}^L \sim P(\mathcal{Y})$ is a binary vector (representing the flattened ICD-9 hierarchy) that indicates whether a label is present or not. We want to minimize the following risk which is the inverse of recall:

$$\mathcal{R}(S) = \mathbb{E}_{X,Y}\left[\sum_{j=1}^{|\mathcal{Y}|} \mathbb{1}\left[Y_j = 1, Y_j \notin S(X)\right]\right]$$

We add **two budget constraints** (1a, 1b) to prevent minimizing the risk by returning all labels and a **hierarchical constraint** (2):
1a) Between K' and K codes are returned per document:

$$\forall_x \in \mathcal{X},\ K' \leq |S(x)| \leq K$$

1b) On average K'' codes are returned over all documents: $\mathbb{E}_X[|S(X)|] \leq K''$
2) Hierarchical constraint: If a leaf node is returned then all of its parent nodes must also be returned:

$$\forall_x \in \mathcal{X}, \forall_y \in \mathcal{Y}, \forall_{\tilde{y}} \in ancestors(y),\ y \in S(x) \Rightarrow \tilde{y} \in S(x)$$

Our goal is to construct a function $S: \mathcal{X} \rightarrow P(\mathcal{Y})$ that satisfies some combinations of the aforementioned constraints. In this paper, we test three different combinations of these constraints: 1) Top-*K* (1a and 2), 2) Average-*K* (1b and 2), and 3) Hybrid (1a, 1b and 2), which is an Average-*K* method with both an upper and a lower bound to prevent returning too many or too few codes.

3. Experiments and Results

Dataset: MIMIC-III [7] is a freely available clinical database. Most studies use the two splits created by Mullenbach et al. [2]. The first split contains the 50 most frequent ICD-9 codes (11,368 records) and the second contains all 8,929 ICD-9 codes (52,722 records) (see Table \ref{tab:donnees}). The code distribution is considerably unbalanced. For example, code *567.2* occurs 211 times, while code *276.5* occurs 1,294 times, or six times as often. Furthermore, in the current splits, there is no guarantee that all the codes will be found in the learning, validation, and test sets. For example, the code *276.5* appears 1,293 times in the learning set, once in the testing set and does not appear at all in the

validation set. We therefore decided to make our own split using the 1,000 most frequent codes in the dataset. We applied a stratification algorithm [8] to ensure that each code was represented in the same proportions in the training, test and validation sets. We also ensured that patients appearing in the training set do not appear in the test/validation sets. The final distribution of the data is presented in Table 1 (MIMIC-III-1000).

Table 1. Statistics on the MIMIC-III splits

Split	Train	Val	Test	Total
MIMIC-III-50	8,066	1,573	1,729	11,368
MIMIC-III-1000[2]	44,592	2,716	5,327	52,635
MIMIC-III-Full	47,719	1,631	3,372	52,722

Implementation: We built an estimator of probabilities $\hat{\eta}$ that uses a neural network on the nodes at the lowest level of the hierarchy and we estimate the parent probabilities on the basis of children scores. We chose LAAT [6] with the optimal parameters mentioned in their paper. We trained the model with a learning rate of 0.001 and a batch size of 8 for 50 epochs. We used early stopping by monitoring micro F1; if there was no improvement after 5 consecutive epochs, we stopped the training. We used the word2vec[3] embeddings trained on all the discharge summaries and a dropout of 0.3 between the embedding and LSTM layers. For text preprocessing, we removed all tokens not containing alphabetic characters and we lowercased all the text. Once the estimator was built, we used it with the Top-*K*, Average-*K* and Hybrid rules.

Evaluation: We evaluated our methods on all the splits but we only present here the results on the MIMIC-III-1000 stratified split. We also evaluated LAAT [6] on our split as our baseline and used the hierarchical recall as our metric. We drew curves that show the trade-off between budget size and micro recall and tested the 3 budget methods on two configurations: 1) taking only leaves from the hierarchy and 2) enriching the labels with their parents.

Analysis and interpretation: Figure 1 shows the results for the different budget approaches. On the ordinate, we have the recall, on the abscissa we have the size of the budget. The red line represents the recall obtained by the baseline, LAAT [6]. The green line shows the average number of labels in the dataset. **The graph on the left** shows the results on the leaves of the hierarchy. Both methods obtain better results than the baseline. For Top-*K*, starting with a budget of 14 (the average number of labels per document), the method improves on the baseline. For Average-*K*, starting with a budget of 12, the method improves on the baseline. On average, Average-*K* achieves a recall that is 5.6% greater than Top-*K*. **The graph on the right** shows the results when using the hierarchy (enriching the labels with parent information). We have removed the leaves that have no siblings and their parents. To be able to compare to a method that does not take the hierarchy into account, we drew curves for when the entire budget is used on the leaves, the parents associated with the selected leaves being added manually. Starting with a budget of 24 (the average number of labels per document), Average-*K* improves on the baseline. For Top-*K*, we need a budget of at least 27 to improve on the baseline. On

[2] Splits created for this publication. https://github.com/leo90v/MIMIC-1000
[3] https://github.com/aehrc/LAAT/tree/master/data/embeddings

average, Average-K achieves a recall 5.61% greater than Top-K. Therefore, allocating the budget to the hierarchy provides a very small improvement in results.

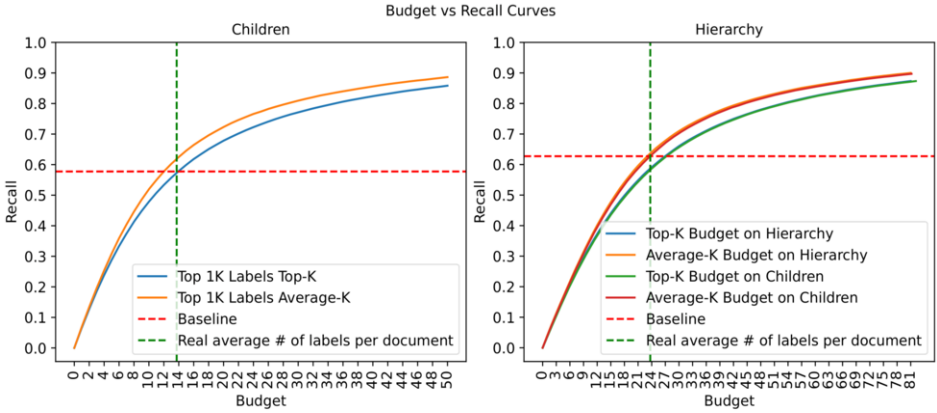


Figure 1. Top-K vs Average-K. In the left graph, we only consider leaf nodes. In the right graph, we consider the hierarchy.

Finally, the results of the hybrid method are shown in Figure 2. Since the hierarchy did not improve the results, we tested this method only on the leaves. We drew the Top-K and Average-K curves as well since they represent the minimum and maximum performance we can obtain with the hybrid method. We set the upper bound (K) as a function of the budget (K''). **In the graph on the left**, we show the micro recall. With $1.2K''$, the performance is on average 3.2% better than with Top-K and 2.25% worse than with Average-K. We also tested with $1.8K''$, but did not draw it since the hybrid and Average-K curves overlap at that point. These results show that with relatively small bounds, the performance is similar to Average-K, while also preventing having an unmanageable number of predictions for some documents. **In the graph on the right**, we show the macro recall. We notice similar behavior but with considerably lower performance. For example, with $1.5K''$ the micro recall is 22% greater on average than the macro recall. This is mainly due to the imbalance of the data set. Even with stratified data, the model has difficulties with some labels.

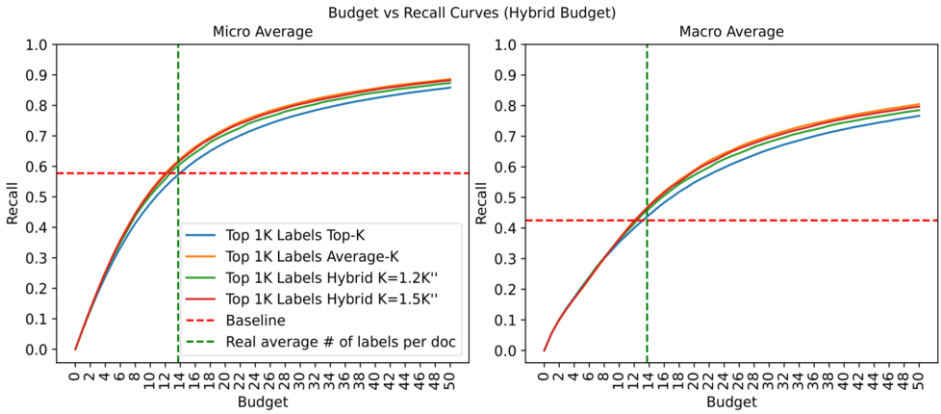


Figure 2. Hybrid method (micro recall on the left and macro recall on the right).

4. Conclusion

We have presented a new approach for semi-automatic clinical coding. We have tested our approach on the splits of the literature and on our stratified split. Via an adaptive predictor, we predict more or less codes per document and at different levels of the hierarchy. Our solution is applicable to any classifier. We plan to use other models in the future, such as LAAT trained with the LDAM loss function [9] designed for unbalanced datasets. We could also use transformers (e.g. Longformer), adapted to long documents [10], but they have not yet surpassed the state of the art for this task. We plan to develop approaches with a weighted budget to give more importance to certain codes. Finally, we plan to experiment our approach using ICD10 and more contemporary data from CHU Montpellier.

References

[1] Richman BD, Kaplan R, Kohli J, Purcell D, Shah M, Bonfrer I, et al. Billing And Insurance-Related Administrative Costs: A Cross-National Analysis. Health affairs. 2022;41 8:1098-106.

[2] Mullenbach J, Wiegreffe S, Duke J, Sun J, Eisenstein J. Explainable Prediction of Medical Codes from Clinical Text. In: 2018 Chapter of the ACL: Human Language Technologies, Volume 1; 2018. p. 1101-11.

[3] Shi H, Xie P, Hu Z, Zhang M, Xing EP. Towards Automated ICD Coding Using Deep Learning. CoRR. 2017;abs/1711.04075.

[4] Teng F, Yang W, Chen L, Huang L, Xu Q. Explainable Prediction of Medical Codes With Knowledge Graphs. Frontiers in Bioengineering and Biotechnology. 2020;8.

[5] Pascual D, Luck S, Wattenhofer R. Towards BERT-based Automatic ICD Coding: Limitations and Opportunities. In: Proceedings of the 20th Workshop on Biomedical Language Processing. Online; 2021. p. 54-63.

[6] Vu T, Nguyen DQ, Nguyen A. A Label Attention Model for ICD Coding from Clinical Text. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20; 2020. p. 3335-41. Main track.

[7] Johnson A, Pollard T, Shen L, Lehman Lw, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific Data. 2016 05;3:160035.

[8] Sechidis K, Tsoumakas G, Vlahavas I. On the Stratification of Multi-label Data. In: Gunopulos D, Hofmann T, Malerba D, Vazirgiannis M, editors. Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg; 2011. p. 145-58.

[9] Cao K, Wei C, Gaidon A, Arechiga N, Ma T. Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019. p. 1567–1578.

[10] Beltagy I, Peters ME, Cohan A. Longformer: The Long-Document Transformer. ArXiv. 2020;abs/2004.05150.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230265

User-Centered Design of a Speech-Based Application to Support Caregivers

Daniel REICHENPFADER[a,1] and Matthias GIRA[b]
[a]*Institute of Business Informatics and Data Science, FH JOANNEUM, Graz, Austria*
[b]*Solgenium OG, Linz, Austria*
ORCiD ID: Daniel Reichenpfader https://orcid.org/0000-0002-8052-3359

Abstract. The shortage of skilled nursing personnel is - among other reasons - due to the low attractiveness of the profession, comprising high workloads and atypical working hours. Studies show that speech-based documentation systems increase documentation efficiency and satisfaction of physicians. This paper describes the development process of a speech-based application to support nurses, according to the user-centered design approach. User requirements were collected based on interviews (n=6) as well as observations (n=6) in three institutions and were evaluated by means of qualitative content analysis. A prototype of the derived system architecture was implemented. Based on a usability test (n=3), further potentials for improvement were determined. The resulting application enables nurses to dictate personal notes, share them with colleagues and transmit notes to the existing documentation system. We conclude that the user-centered approach ensures the extensive consideration of the nursing staff's requirements and shall be continued for further development.

Keywords. eHealth, mobile applications, nursing documentation, User-Centered Design, speech recognition

1. Introduction

A study on the attractiveness of the nursing profession in Austria shows that 65 % of respondents consider it unlikely that they will continue in their profession until retirement. 15 % say they already have concrete intentions or plans to change jobs. The main reasons given for this low attractiveness is the high workload, time pressure and emotional stress [1]. For patients or residents of nursing homes, this shortage of caregivers in combination with the resulting overload is fatal:

According to the Ombudsman Board Austria, human rights were violated in nursing homes: Due to understaffing, beds were blocked as well as activities with residents and access to the outdoors restricted. These examples demonstrate that improvements in staffing resources are needed to ensure the quality of life of residents [2].

In addition to the shortage of personnel, the increasing amount of nursing documentation is also considered a burden: Regarding inpatient nursing, up to 35 % of total working time is spent on documentation tasks [3]. Increasing digitization in the healthcare sector is not providing any relief: electronic medical records lead to increased documentation quality, but also to greater documentation effort [4].

[1] Corresponding Author: Daniel Reichenpfader, E-mail: hello@danielreichenpfader.com.

One starting point for improving documentation by means of information and communications technology (ICT) could be the application of speech-based systems in nursing. Speech-based documentation (speaking into a microphone rather than typing out the text) is three or up to ten times faster than typing or writing by hand, respectively [5]. Vogel et al. show that speech-based documentation increases user satisfaction by 23 %, documentation speed by 26 % and the extent of documentation content by 82 % [6]. However, the last aspect can lead to increased effort for searching for information. Therefore, within the scope of this work, a speech-based system for supporting nursing documentation is developed. We focused on integrating caregivers into the entire development process according to the user-centered design (UCD) paradigm. This inclusion facilitates the ascertainment of the staff's requirements as comprehensively and correctly as possible.

2. Methods

As a first step, we conducted six guideline-based interviews and six structured observations with nurses in three Austrian healthcare institutions in order to identify user requirements. The interview transcripts and observation protocols served as input to a qualitative content analysis according to Mayring, which we performed using the software tool MAXQDA Plus 2022 [7]. For this analysis, we defined the following three research questions, providing guiding principles for the subsequent deduction of analysis categories:

- How is nursing documentation - regarding processes, procedures and contents - organized in the institution?
- What are the most severe problems experienced subjectively by the caregivers regarding nursing documentation?
- What is the attitude of nurses towards the use of speech-based mobile applications?

Utterances and observations were then classified into one of these categories. Paraphrases of these categorized elements in turn enabled the derivation of system requirements for the platform and the design of a system architecture. Next, we iteratively developed a prototype application of the system. iOS was chosen as target platform due to existing programming experience of the authors, as well as simple integration of built-in speech-recognition capabilities (Speech framework). After each iteration, a feedback interview was carried out with a nurse, in order to get direct feedback and to identify potentials for improvement. In order to transfer data from the mobile device to the legacy nursing documentation system, we developed a Java application to be installed on the hospital workstation. For the backend, we used a Software-as-a-Service platform (Google Firebase) to implement basic user management and a database to store documentation entries. Last, a concluding usability test was conducted with three different nurses, having three, ten and 30 years of professional experience respectively. Each person had to fulfil three tasks, comprising creation, sharing and transmission of documentation entries.

3. Results

The six conducted interviews lasted between 15 and 30 minutes and an average of 24 minutes. Interview transcripts and observation protocols formed the corpus for the subsequent qualitative content analysis. We identified and assigned 333 text passages to one of 15 deductively defined categories. The paraphrased passages served as the basis for defining 30 system requirements of the application to be developed.

There exist three common aspects in all three institutions, which became apparent during the qualitative content analysis. First, each institution maintains an ongoing report on the condition of a patient or resident. This is referred to as a nursing report. In some institutions, other professional groups also contribute entries in this report, which therefore serves as a main communication hub between professions, colleagues and shifts.

The second common aspect is the creation of personal notes by nursing staff: In order to avoid waiting times when starting up workstations and logging in and out, nurses usually jot down pending and completed tasks, important information about patients, and reminders. However, this information typically stays within the nurse's pocket on a piece of paper and is not shared.

The third common aspect is the documentation timing: In all three institutions, documentation is entered into a hospital information system or paper-based patient record for several patients or residents at once and only after the routine activities have been performed. Nurses state that these delayed and batchwise documentation practices are error prone.

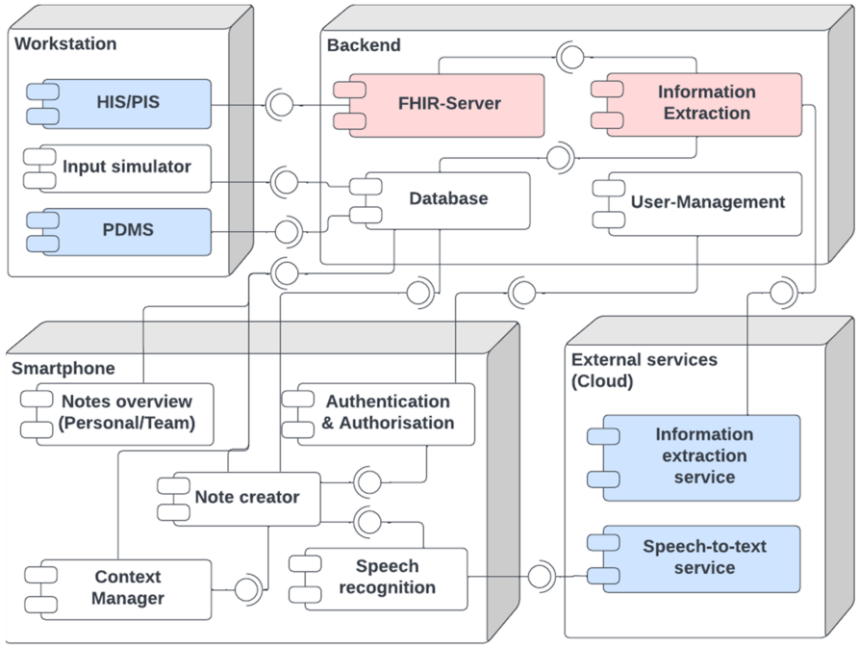


Figure 1. UML component diagram of the system architecture (blue = components provided by external systems, red = components not implemented as prototype).

Based on these three common aspects, the system architecture shown in Figure 1 was designed. The developed prototype provides the following functionalities: Nursing report entries as well as personal notes are created using speech-recognition with a single tap, shared with the team and marked as completed. A context manager component enables the quick and secure identification of the patient to be documented, e.g. by scanning a barcode or NFC tag. For speech recognition, the Speech framework provided by iOS was used. The developed Java application (Input simulator) enables immediate pasting of notes into the currently focused text field at the workstation upon a single tap on the mobile device. Implementation is based on the java.awt.datatransfer package.

During prototype development, two iterations were completed. The first feedback interview enabled us to identify 15 usability problems, of which nine were fixed. Respectively, we fixed three out of twelve usability problems after the second development iteration. Based on the concluding usability test, five additional potentials for improvement were identified.

4. Discussion and Conclusions

The proposed system architecture facilitates the creation of nursing report entries, replaces paper-based handwritten personal notes of nurses and allows for immediate documentation, one patient at a time. We chose these three aspects to be the core functionalities of the designed system as we assume that they might be generalizable to other Austrian healthcare institutions, therefore augmenting the target group of potential users.

As part of the prototype, the developed Java application enables the transmission and insertion of notes to the existing electronic documentation system on Windows-based systems via simulating input from the system clipboard. This low-level approach ensures vendor-independent interoperability. In future, structured information might as well be transmitted via standardized interfaces and exchange formats, handled by the *FHIR-Server* component. The proposed system is not intended to replace existing speech-recognition tools, but rather to provide an intuitive interface to use these services, regardless of the manufacturer. While for the prototype, on-device speech-recognition was used, integration of external speech-recognition services is already considered in the system architecture.

Latif et al. show that the application of speech-based documentation has high potential to improve different healthcare settings [8]. The authors point out that one challenge is to ensure interoperability between systems, which we considered in our system architecture. Within the nursing domain, speech-based documentation has recently gained momentum with similar software systems being currently developed [9, 10]. In contrast to existing systems however, our proposed solution provides vendor-independent interoperability on two levels: Documentation entries can either be pasted into an arbitrary text field (low-level) or transmitted as standardized and structured FHIR resources. While the low-level approach does not require any changes to existing documentation systems, the support of FHIR ensures future interoperability as the adoption of this standard is continuously advancing among information system vendors. Furthermore, we intentionally followed the UCD approach to acquire a deep understanding of nurses' challenges regarding documentation in day-to-day operations. This methodology enabled us to design a system fitted to the needs of the profession, generalized to three – and possibly more – Austrian healthcare institutions.

Although this study yielded valuable insights, there are limitations that must be acknowledged: Due to the relatively small sample size of six interviewees and six observation participants, transferability to other institutions can only be hypothesized. Moreover, due to time and resource constraints, we decided to omit implementation of the system components regarding automatic information extraction; these components however are essential to enable the fully automated extraction of speech-based documentation entries, omitting the need to transfer entries to the documentation system at the workstation. Furthermore, we developed the application prototype for the iOS platform only, which some test persons might not have been familiar with. Nevertheless, we put focus on the development of a simple and easily understandable mobile user interface for this platform. Usability testing was facilitated by implementing the remaining system components based on a Minimum Viable Product (MVP) approach.

Additional functionalities such as support for care documentation in the native language of a caregiver and automatic translation could be similarly investigated for their acceptance based on the described UCD approach. Conducting quantitative research methods based on the fully implemented system, e.g. by introducing and testing the system at a hospital ward, could provide statistical evidence of the system's impact on nurse productivity and satisfaction as well as efficiency and safety of nursing documentation. In the future, our system can contribute to the improvement of working conditions in nursing as well as to increasing the attractiveness of nursing professions and thus counteract the acute shortage of nursing staff.

References

[1] Federal Ministry of Social Affairs, Health, Long-Term Care and Consumer Protection Austria (BMSGPK). Arbeitsbedingungen in Pflegeberufen. Sonderauswertung des Österreichischen Arbeitsklima Index [Working Conditions in Nursing Professions. Special Evaluation of the Austrian Work Climate Index]. 2021.

[2] Austrian Ombudsman Board. Bericht der Volksanwaltschaft an den Nationalrat und an den Bundesrat 2021 [Report of the Ombudsman Board to the National Council and the Federal Council 2021]. 2022. German.

[3] Collins S, Couture B, Kang MJ, Dykes P, Schnock K, Knaplund C, Chang F, Cato K. Quantifying and Visualizing Nursing Flowsheet Documentation Burden in Acute and Critical Care. AMIA Annu Symp Proc. 2018 Dec;2018:348-357. PMID: 30815074; PMCID: PMC6371331.

[4] Baumann LA, Baker J, Elshaug AG. The impact of electronic health record systems on clinical documentation times: A systematic review. Health Policy. 2018 Aug;122(8):827–36. doi: 10.1016/j.healthpol.2018.05.014

[5] Kumah-Crystal Y, Pirtle C, Whyte H, Goode E, Anders S, Lehmann C. Electronic Health Record Interactions through Voice: A Review. Appl Clin Inform. 2018 Jul;09(03):541–52. doi: 10.1055/s-0038-1666844

[6] Vogel M, Kaisers W, Wassmuth R, Mayatepek E. Analysis of Documentation Speed Using Web-Base Medical Speech Recognition Technology: Randomized Controlled Trial. J Med Internet Res. 2015 Nov 3;17(11):e247. doi: 10.2196/jmir.5072

[7] Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken [Qualitative content analysis: Foundations and methods]. Weinheim Basel: Beltz; 2015. German.

[8] Latif S, Qadir J, Qayyum A, Usama M, Younis S. Speech Technology for Healthcare: Opportunities, Challenges, and State of the Art. IEEE Rev Biomed Eng. 2021;14:342–56. doi: 10.1109/RBME.2020.3006860

[9] Schmidberger M. Voize [Internet]. Potsdam: Voize GmbH; c2022 [cited 2023 Mar 2]. Available from: https://www.voize.de

[10] Nuance Communications, Inc. AI Powered Clinical Documentation Solutions for Nurses [Internet]. [cited 2023 Mar 2]. Available from: https://www.nuance.com/en-gb/healthcare/care-settings-specialties/nursing.html

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230266

Investigating Canadian Public Attitudes Toward COVID-19 Vaccine Mandates with a Nested Analysis Framework

Yang YANG[a,1], Xiaomeng LEI[b], Zeping ZHOU[b], Shu-Feng TSAO[a], Zahid A. BUTT[a] and Helen H. CHEN[a]
[a]*School of Public Health Sciences, Canada*
[b]*Computer Science, University of Waterloo, Canada*

Abstract. Background: Social media is an important medium for studying public attitudes toward COVID-19 vaccine mandates in Canada, and Reddit network communities are a good source for this. **Methods**: This study applied a "nested analysis" framework. We collected 20378 Reddit comments via the Pushshift API and developed a BERT-based binary classification model to screen for relevance to COVID-19 vaccine mandates. We then used a Guided Latent Dirichlet Allocation (LDA) model on relevant comments to extract key topics and assign each comment to its most relevant topic. **Results**: There were 3179 (15.6%) relevant and 17199 (84.4%) irrelevant comments. Our BERT-based model achieved 91% accuracy trained with 300 Reddit comments after 60 epochs. The Guided LDA model had an optimal coherence score of 0.471 with four topics: travel, government, certification, and institutions. Human evaluation of the Guided LDA model showed an 83% accuracy in assigning samples to their topic groups. **Conclusion**: We develop a screening tool for filtering and analyzing Reddit comments on COVID-19 vaccine mandates through topic modelling. Future research could develop more effective seed word-choosing and evaluation methods to reduce the need for human judgment.

Keywords. COVID-19, Canada, vaccine mandate, public attitudes, Reddit, social media, topic modelling, BERT, Guided LDA

1. Introduction

Social media data has been used to study public health issues and attitudes toward vaccination [1]. Reddit is an important communication channel, and its volume of data grows with web technology advances. To analyze large amounts of text data, it is necessary to develop methods for classifying and filtering this "Big Data," often using NLP techniques. Zhu *et al.* [2] developed a pre-training method based on the Bidirectional Encoder Representations from Transformers (BERT) model to classify Tweets about personal experiences with medication. Klein *et al.* [3] proposed using the BERT model to identify potential COVID-19 cases by collecting and filtering tweets with self-reported symptoms and geolocations. Muller, Salathé, and Kummervold [4] developed CT-BERT, a BERT model optimized for COVID-19 tasks by being pre-trained on Twitter data related to the pandemic.

[1] Corresponding Author: Yang Yang, E-mail: y24yang@uwaterloo.ca

Latent Dirichlet Allocation (LDA) is a probabilistic model for learning the structure of discrete data collections, such as corpora [5,6]. It is frequently used to discover topics and analyze social media data. LDA has been applied to Twitter data, achieving an accuracy of 98% in classifying topics [7]. Researchers have used LDA to analyze the themes and patterns present in tweets and examine the relationships between those themes [8]. The Guided LDA model is a variant of LDA that addresses limitations such as topic overlapping and meaningless topics. It allows the specification of seed words to guide the model to focus on certain terms and has an *eta* parameter that tracks word assignment to topics. However, using a prior distribution with seed words can influence the model to concentrate on specific topics. This project aimed to analyze Canadian public discourse about COVID-19 vaccine mandate policies announced in September 2021 by applying a "nested analysis" framework and topic modelling techniques to Reddit data collected in Canada.

2. Methods

2.1. Related Work

Researchers have proposed various pre-training and fine-tuning methods based on BERT and RoBERT models [9] for COVID-19-related tasks. Sentence-BERT (SBERT) [10] was introduced to compare semantically meaningful sentence embeddings, and BioBERT [11] was proposed for analyzing biomedical documents. Latent Semantic Analysis (LSA) [12] and LDA are mathematical methods that have been applied to topic modelling and sentiment analysis of Twitter data. LDA has shown better accuracy than LSA in topic modelling [7] and is useful in identifying tweet patterns, themes, and structures [8].

2.2. Dataset and Preprocessing

We collected 20378 Reddit comments from Canada in September 2021 and labelled each comment as relevant or irrelevant to COVID-19 vaccine mandates. Of the total samples, 17199 (84.4%) comments were irrelevant, and 3179 (15.6%) were relevant comments. An under-sampling strategy was employed to create a balanced training set α with an equal amount of 3179 relevant and irrelevant comments. The data preprocessing steps comprised word tokenization, stopword removal, word stemming, and synonyms merging.

2.3. Binary Classification Using BERT-based Model

The BERT-based binary classifier with DistilRoBERTa was trained using different amounts of data, ranging between 100 and 1000 comments randomly selected from α. The BERT-based model maps sentences to a 768-dimensional vector space, producing a feature vector for each comment and using a neural network with two fully connected layers and an output layer with Relu, Relu, and Softmax activation functions for binary classification.

2.4. Topic Modelling Using Guided LDA

We initially tried the standard LDA model to explore vaccine mandate topics, but it struggled to split themes. We then developed a Guided LDA model, which selects seed words related to each topic based on the standard LDA model results to improve performance.

3. Results

3.1. Binary Classification Using BERT-based Model

The BERT-based model performed best with training on 300 comments randomly selected from α, achieving 91% accuracy on the entire dataset after 60 epochs (Table 1).

Table 1. BERT-based binary classification model performance with different sizes of datasets.

Size of Dataset (# of Reddit comments)	Accuracy	Convergence Speed (Epochs)
100	0.83	275
150	0.86	150
200	0.89	100
300	**0.91**	60
400	0.90	60
500	0.90	40
1000	0.89	25

The loss of the training set decreased from 0.7 to 0.3, and the accuracy increased to approximately 0.9 (Figure 1).

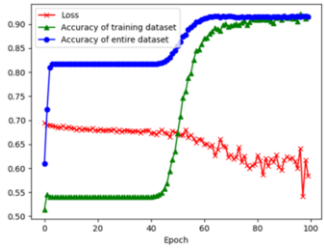


Figure 1. The BERT-based binary classification model trained on 300 Reddit comments.

3.2. Topic Modelling Using Guided LDA

The Guided LDA model divided the Reddit comments into four topics based on the optimal coherence score of 0.47 (Figure 2). The model's performance was evaluated by manual interpretation, which showed that 83% of the samples were accurately assigned to their topic group. Table 2 summarizes the four topics and corresponding seed words.

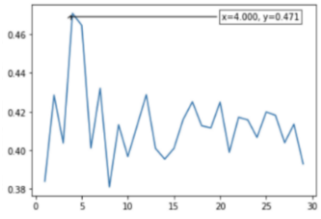


Figure 2. Coherence scores of Guided LDA model for different topic numbers. X-axis: number of topics; Y-axis: coherence score

Table 2. Seed words for Guided LDA.

Topic	Seed Words
Travel	"test" "travel" "day" "border" "result" "hour" "quarantine" "pcr" "time" "flight" "trip" "plan" "home" "country" "week" "return" "airport" "enter" "hotel"
Government	"govern" "law" "right" "province" "policy" "court" "gov" "protest" "employ" "vote" "leadership" "party" "elect" "win" "province" "mask" "rule" "service"
Certification	"code" "card" "qr" "passport" "health" "system" "app" "id" "proof" "receipt" "mandate" "certificate" "restrict"
Institution	"work" "employer" "employee" "hospital" "business" "school" "kid" "number" "case" "rate"

4. Discussion

This study aimed to understand the public attitudes toward COVID-19 vaccine mandates in Canada by analyzing comments on the Reddit platform using a nested analysis framework. We developed a pipeline that utilizes a BERT-based model for binary classification of Reddit comments related to COVID-19 vaccine mandates and a Guided LDA model to further analyze the topics in the relevant comments. Our study found that Canadian public discourse on COVID-19 vaccine mandates revolved around four main themes: travel, government, certification, and institutions. These themes encompassed discussions on the impact of vaccine requirements on travel [13], the government's role in implementing and enforcing mandates [14,15], proof of vaccination and vaccine passports [16], and the role of schools, workplaces, health care, and other institutions in implementing mandates [14,15].

The BERT-based model achieved an accuracy of 91% with a small training set of 300 Reddit comments and converged faster with larger training sets, demonstrating its usefulness in identifying relevant Reddit comments related to COVID-19 vaccine mandates. Our results also showed that having more training data did not always lead to better performance, as the model's accuracy dropped slightly to 89% when trained on 1000 comments. The Guided LDA model achieved 83% accuracy in human evaluation, improving standard LDA through careful tuning with seed words for each topic. This allowed for a more accurate understanding of public attitudes toward COVID-19 vaccine mandates from a large volume of Reddit comments. Still, the quality of the seed words and evaluation limitations may affect the model's performance and require manual interpretation.

This study applies a "nested analysis" framework with NLP topic modelling and human judgement for model evaluation and iterative data preprocessing. The Pushshift API allowed for efficient analysis of large quantities of Reddit data over a specific time period with greater flexibility and access to additional parameters, making it superior to keyword searches with the Reddit search engine. This approach allowed for discovering emerging topics and identifying patterns and themes using topic modelling techniques such as LDA. One limitation is that Reddit comments may not represent the overall Canadian population's views on COVID-19 vaccine mandates and may not be reliable sources of information. Future research should consider alternative methods for classifying datasets, evaluating the Guided LDA model, and selecting seed words to improve the model's performance.

5. Conclusion

We use NLP techniques on Reddit comments to investigate Canadian attitudes toward COVID-19 vaccine mandates and demonstrate their usefulness for health researchers. Filtering social media data and focusing on specific aspects, especially when dealing with large amounts of text data, can improve understanding of public health issues and inform effective interventions.

References

[1] Müller MM, Salathé M. Crowdbreaks: tracking health trends using public social media data and crowdsourcing. Frontiers in public health. 2019 Apr 12;7:81.
[2] Zhu M, Song Y, Jin G, Jiang K. Identifying personal experience tweets of medication effects using pre-trained Roberta language model and its updating. In Proceedings of the 11th international workshop on health text mining and information analysis 2020 Nov (pp. 127-137).
[3] Klein AZ, Magge A, O'Connor K, Flores Amaro JI, Weissenbacher D, Gonzalez Hernandez G. Toward using Twitter for tracking COVID-19: a natural language processing pipeline and exploratory data set. Journal of medical Internet research. 2021 Jan 22;23(1):e25314.
[4] Müller M, Salathé M, Kummervold PE. Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. arXiv preprint arXiv:2005.07503. 2020 May 15.
[5] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun 1;155(2):945-59.
[6] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research. 2003;3(Jan):993-1022.
[7] Negara ES, Triadi D, Andryani R. Topic modelling twitter data with latent dirichlet allocation method. In2019 International Conference on Electrical Engineering and Computer Science (ICECOS) 2019 Oct 2 (pp. 386-390). IEEE.
[8] Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PloS one. 2020 Sep 25;15(9):e0239441.
[9] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692. 2019 Jul 26.
[10] Reimers N, Gurevych I. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084. 2019 Aug 27.
[11] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020 Feb 15;36(4):1234-40.
[12] Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Discourse processes. 1998 Jan 1;25(2-3):259-84.
[13] Charrier L, Garlasco J, Thomas R, Gardois P, Bo M, Zotti CM. An Overview of Strategies to Improve Vaccination Compliance before and during the COVID-19 Pandemic. International Journal of Environmental Research and Public Health. 2022 Sep 3;19(17):11044.
[14] Griffith J, Marani H, Monkman H. COVID-19 vaccine hesitancy in Canada: content analysis of tweets using the theoretical domains framework. Journal of medical Internet research. 2021 Apr 13;23(4):e26874.
[15] Benham JL, Atabati O, Oxoby RJ, Mourali M, Shaffer B, Sheikh H, Boucher JC, Constantinescu C, Parsons Leigh J, Ivers NM, Ratzan SC. COVID-19 vaccine–related attitudes and beliefs in Canada: National cross-sectional survey and cluster analysis. JMIR Public Health and Surveillance. 2021 Dec 23;7(12):e30424.
[16] Flores Anato JL, Ma H, Hamilton MA, Xia Y, Harper S, Buckeridge D, Brisson M, Hillmer MP, Malikov KA, Kerem A, Beall R. Impact of a vaccine passport on first-dose COVID-19 vaccine coverage by age and area-level social determinants in the Canadian provinces of Québec and Ontario: an interrupted time series analysis. medRxiv. 2022:2022-10.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230267

Secondary Use of Clinical Problem List Entries for Neural Network-Based Disease Code Assignment

Markus KREUZTHALER[a,1], Bastian PFEIFER[a], Diether KRAMER[b] and Stefan SCHULZ[a]
[a] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*
[b] *Department of Information and Process Management, Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria*

Abstract. Clinical information systems have become large repositories for semi-structured and partly annotated electronic health record data, which have reached a critical mass that makes them interesting for supervised data-driven neural network approaches. We explored automated coding of 50 character long clinical problem list entries using the International Classification of Diseases (ICD-10) and evaluated three different types of network architectures on the top 100 ICD-10 three-digit codes. A fastText baseline reached a macro-averaged F_1-score of 0.83, followed by a character-level LSTM with a macro-averaged F_1-score of 0.84. The top performing approach used a downstreamed RoBERTa model with a custom language model, yielding a macro-averaged F_1-score of 0.88. A neural network activation analysis together with an investigation of the false positives and false negatives unveiled inconsistent manual coding as a main limiting factor.

Keywords. Natural Language Processing, Electronic Health Records, Machine Learning, Secondary Use

1. Introduction and Motivation

The clinical information system (CIS) of a large, multicentre public hospital provider in Austria stores short clinical problem descriptions in German (maximum 50 characters) together with which manually assigned codes from the International Classification of Diseases (ICD-10). This huge table fulfils three purposes: (i) collection of content that automatically fills the "Diagnoses" section when narrative discharge summaries are created, (ii) display of a problem-list like scrollable textbox in the CIS frontend, and (iii) provision of ICD codes for administrative purposes.

Due to the technical limitation of 50 characters, the often-lengthy ICD-10 descriptions are usually overwritten by the users, who use overly compact expressions, characterized by ellipsis, context-dependent abbreviations and acronyms, non-standardised numeric values, spelling variants and errors. These text snippets exemplify typical idiosyncrasies of clinical language [1].

[1] Corresponding Author: Markus Kreuzthaler, E-mail: markus.kreuzthaler@medunigraz.at

Our work is centred on this resource. We wanted to investigate to what extent clinical real-world data is suited as training material for different types of multi-class classification approaches for automatic assignment of codes from the classification system ICD-10, and where it reaches its limits. We applied three different types of neural network (NN) architectures: a shallow NN, a recurrent NN and a transformer-based architecture for our experimental secondary use-case scenario of clinical routine documentation.

2. Methods and Data

We used ~1.9 million unique de-identified problem list entries and ignored all entries without ICD codes. A 90/10 split was carried out for training and test set preparation. We trained one model using the top 100 occurring three-digit ICD-10 codes, which cover about 50% of all code instances in the data set. Therefore, we scaled up all training samples of this highly imbalanced data set to the most frequent code, resulting in ~6 million training samples (100 times ~60k observations). The test set based distribution of the 100 codes under investigation remained unchanged (~93k observations).

2.1. Neural Network Architectures

Shallow. As a baseline, we used *fastText* [2], exploiting pre-trained skip-gram embeddings at subword and word type level from the training set. We used a simple rule-based tokenizer[2] and normalized all resulting tokens to lower case.
Recurrent. Character inputs are modelled as 122 dimensional one-hot encoded vectors from the training set, together with an out-of-character-dictionary feature dimension in accordance to [3] for the input to the chosen LSTM [4] network. *Deeplearning4j* was used as implementation library.
Transformer. For transformer-based architectures, we decided to apply RoBERTa [5]. We build our own language model, in order to support the downstream task with a first understanding of the language under scrutiny. To this end, we used *ktrain*, a lightweight wrapper library, for *Hugging Face*.

2.2. Neural Network Interpretation

In health care scenarios, regulatory frameworks require that decision support systems are able to explain the path that led to a certain suggestion. This is optimally met by decision trees, but poses complex challenges for NNs. Established approaches are LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations) or the notion of *saliency* as being the norm of the gradient of the loss function to a given input, an approach successfully applied in clinical NLP domains [6] for explainable machine learning systems. In this work, we were particularly interested in character-wise feedback for the overall classification result via the inspection of certain class probabilities at certain positions of the LSTM sequence model. This allows the identification of activation levels at this granular input representation scheme, because single characters have significant impact on the correct interpretation of narrative content.

[2] [^\p{IsAlphabetic}\p{IsDigit}]

2.3. Evaluation Metrics

We evaluated the performance of the trained model on the test data set using precision, recall, and F_1-score per ICD-10 code, as well as macro-evaluation statistics. The following definitions were used: true positives (*TPs*) – the number of correctly assigned codes; true negatives (*TNs*) – the number of correctly unassigned codes; false positives (*FPs*) – the number of incorrectly assigned codes; false negatives (*FNs*) – the number of incorrectly unassigned codes. Exploiting this definition, precision $P = TP / (TP + FP)$, recall $R = TP / (TP + FN)$ and $F_1\text{-}score = 2 \cdot P \cdot R / (P + R)$.

3. Results and Discussion

Table 1. Selection of test set based evaluation results using RoBERTa.

ICD-10	Precision	Recall	F_1-score	ICD-10	Precision	Recall	F_1-score
P07	0,99	1,00	0,99	D48	0,79	0,71	0,74
F17	0,99	0,97	0,98	D37	0,74	0,73	0,73
A46	0,96	0,97	0,97	I47	0,67	0,77	0,72
E78	0,96	0,97	0,97	T14	0,83	0,64	0,72
G47	0,97	0,98	0,97	I64	0,56	0,72	0,63
I35	0,97	0,97	0,97	E10	0,53	0,73	0,61
I71	0,95	0,98	0,97	N19	0,43	0,73	0,54
D64	0,95	0,97	0,96	E14	0,45	0,51	0,47
G40	0,96	0,96	0,96	C80	0,40	0,55	0,46
J44	0,96	0,96	0,96	Z03	0,38	0,44	0,41

Evaluation of the test set with 92,832 problem list entries yielded an overall macro-averaged F_1-score of 0.83 for the fastText baseline, 0.84 for the LSTM, and 0.88 for the RoBERTa approach. Table 1 shows the evaluation on an ICD-10 three-digit level. The left side of the table shows the top 10 best performing codes, the right side shows codes with a macro-averaged F_1-score less than 0.75. We got the highest F_1-score value for the ICD-10 code P07 ("disorders related to short gestation and low birth weight, not elsewhere classified") and the lowest one for Z03 ("medical observation and evaluation for suspected diseases and conditions") for all three different NN architectures. Interestingly, LSTM, which solely relies on character-level inputs, showed about the same performance as the fastText baseline. This provides strong evidence that the LSTM network is highly sensitive to even minor variations of character sequences that are relevant for the correct ICD code assignment. We decided to investigate this further by performing a network activation analysis of the LSTM.

3.1. Explainable AI

What can be seen at the top of Figure 1 is the network stimulation for the correct class I25 (chronic ischaemic heart disease) for the German input "Kononare Herzkrankh. 1 Gefäß – 1 x DES in LAD" (coronary heart disease. 1 vessel – 1 x DES in LAD). The intensity level corresponds to the class level probability for a given code at a given position and varies between zero and one. A low-level intensity therefore corresponds to

a low probability for the class shown on the right in the figure. This snippet exhibits the typing mistake “Kononare” instead of “Koronare” (coronary), an ad-hoc abbreviation “Herzkrankh.” for “Herzkrankheit” (heart disease) and two acronyms. “DES” stands for “drug-eluting stent”and “LAD” for “left anterior descending (artery)”. It is principally the last part of this input string that strongly supports the code I25. In contrast, the same string has a very low feedback activation for the ICD-10 class E11 (type 2 diabetes mellitus), as expected.

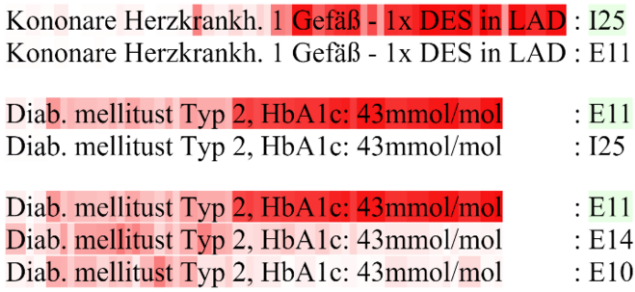


Figure 1. Activation heat map with respect to a given class at a certain character position.

The centre of Figure 1 displays the result of the experiment in reverse. We see the network activation for the input “Diab. mellitust Typ 2, HbA1c: 43 mmol/mol” (diabetes mellitus type 2, HbA1c: 43 mmol/mol) and its corresponding correct code E11. Again, there are ad-hoc abbreviations “Diab." for “Diabetes” and the typing mistake “mellitust” (correct “mellitus”). With the occurrence of the digit “2” in the input string, the network responds with a very high activation at this position for type 2 diabetes mellitus. “HbA1c” (glycated haemoglobin) is also an important diabetes biomarker. Conversely, there is very little feedback activity for the class I25, resulting in low probability values for the whole input sequence.

At the bottom of Figure 1 the input activation for the same string is contrasted with E10 (type 1 diabetes mellitus) and E14 (unspecified diabetes mellitus). As seen before with the appearance of the character “2”, there is strong evidence for E11. Nevertheless, for E10 and E14, there are clearly observable activation levels. This inspection supports our assumption that, for certain ICD-10 code sections, the network will intermix classes if there is no clear consistent exclusive manual coding. Nevertheless, the network tries to interpret the input as well as possible with respect to the training data.

4. Conclusion and Outlook

We presented an approach for the assignment of short clinical problem list entries to ICD-10 three-digit codes using three different NN architectures: a shallow NN (fastText), a recurrent NN (LSTM) and a transformer-based architecture (RoBERTa) which performed best with an overall macro-averaged F_1-score of 0.88. The fastText baseline evaluated with a macro-averaged F_1-score of 0.83, the LSTM approach modelling the problem purely on character-level reached a macro-averaged F_1-score of 0.84. In a character-level based heat map, we visualized LSTM network activations to a given input. Thus, we could trace classification decisions back. TP, FP and FN analyses revealed that the trained network suffered from coding inconsistencies (e.g., E14 versus E11).

The preliminary experiments presented in this paper emphasize the potential of secondary use scenarios of annotated semi-structured clinical real-world data for building NN-based applications such as the assignment of disease codes. Taking advantage of such resources is needed for data-driven applications such as the generation of clinical BERT models [7] for languages other than English, like German, where to the best of the authors knowledge a model is not available by now. There is yet an unexploited potential of NNs to bridge the gap between standardized code entries and the language used in the routine of clinical documentation and communication. Identifying and leveraging annotated language resources as presented in this paper could be a step in this direction.

Acknowledgments

This study was approved by the ethics committee of the Medical University of Graz (30-496 ex 17/18).

References

[1] Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical Natural Language Processing in languages other than English: opportunities and challenges. J Biomed Semant 2018;9:12. https://doi.org/10.1186/s13326-018-0179-8.

[2] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Trans Assoc Comput Linguist 2017;5:135–46. https://doi.org/10.1162/tacl_a_00051.

[3] Zhang X, LeCun Y. Text Understanding from Scratch 2016. https://doi.org/10.48550/arXiv.1502.01710.

[4] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

[5] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach 2019. https://doi.org/10.48550/arXiv.1907.11692.

[6] Gehrmann S, Dernoncourt F, Li Y, Carlson ET, Wu JT, Welt J, et al. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. PLOS ONE 2018;13:e0192360. https://doi.org/10.1371/journal.pone.0192360.

[7] Alsentzer E, Murphy J, Boag W, Weng W-H, Jindi D, Naumann T, et al. Publicly Available Clinical BERT Embeddings. Proc. 2nd Clin. Nat. Lang. Process. Workshop, Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019, p. 72–8. https://doi.org/10.18653/v1/W19-1909.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230268

Ontology-Based Semantic Annotation of French Psychiatric Clinical Documents

Ons AOUINA[a,1], Jacques HILBEY[a] and Jean CHARLET[a,b]
[a] *Sorbonne Université, Sorbonne Paris Nord, INSERM, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé - LIMICS, Paris, France*
[b] *Assistance Publique-Hôpitaux de Paris, Paris, France*

Abstract. Building a timeline of psychiatric patient profiles can answer many valuable questions, such as how important medical events affect the progression of psychosis in patients. However, the majority of text information extraction and semantic annotation tools, as well as domain ontologies, are only available in English and cannot be easily extended to other languages, due to fundamental linguistic differences. In this paper, we describe a semantic annotation system based on an ontology developed in the PsyCARE framework. Our system is being manually evaluated by two annotators on 50 patient discharge summaries, showing promising results.

Keywords. Semantic Annotation, GATE, Ontology, Psychiatry, NLP

1. Introduction

Schizophrenia and chronic psychosis are among the most debilitating disorders in adolescents and young adults and are associated with cognitive impairment, poorer occupational success, and poor quality of life. Studies have shown that the longer the duration of untreated psychosis, the worse the outcome of intervention, the worse the recovery and general functioning, and the greater the long-term social impairment [1]. This issue is addressed by the RHU PsyCARE[2] project, which aims to improve early detection and intervention in psychoses.

In this context, the analysis of Patient Discharge Summaries PDSs can give us the opportunity to study many valuable questions such as how important medical events affect the progression of psychosis in patients. These summaries provide information about the patient's history (e.g., associated with the onset of symptoms or the start of treatment). However, extracting such information to trace the history of psychosis and develop the timeline is a complex matter that requires carefully annotated corpora. Being able to automatically extract this information would improve medical care and support clinical research.

In this project, we propose a method of semantic annotation of PDSs based on an ontology developed in the framework of PsyCARE. Our approach will not only extract medical entities from a text but also transform them into structured and formalized knowledge. Ontologies allow the design of semantic data indexes that leverage medical

[1] Corresponding Author: Ons AOUINA; E-mail: ons.aouina@etu.sorbonne-universite.fr.
[2] https://psy-care.fr/

knowledge to improve information retrieval and search [2]. This proposal is a first brick in a large project that aims to build a complete timeline of a patient's psychosis based on his medical records.

2. Material and Methods

2.1. Psychiatry Dataset

The clinical documents used in this work are derived from the French PsyCARE project. It is a compilation of about 8000 anonymized Patient Discharge Summaries covering a period of ten years, which represents a volume of about 3,500,000 words. These summaries come from the Groupe Hospitalier Universitaire Psychiatrie et Neurosciences de Paris, the largest psychiatric hospital in Paris. They are semi-standardized, in Word format and have been strictly anonymized beforehand, by deleting all names, dates, places, etc. In addition, the diagnosis is indicated at the end of each document and is referenced to the ICD-10[3].

These records are written in French and describe the patient's history and social context, medications, details of hospital admission, and current and previous psychiatric diagnoses.

2.2. Domain Ontology description

The ontology developed in the framework of PsyCARE both to integrate the data and to allow their semantic annotation represents domains such as psychiatric clinical aspects, drugs with their ATC code, imaging, biology, etc. In our context we are interested in the psychiatric clinical branch which describes the signs, symptoms, disorders related to psychiatry as well as the medication branch.

Based on this ontology, we reconstructed an annotation scheme to which we added a branch describing the structure of the PDSs[4]. This allows us to relate concepts to their context of occurrence. For example, drugs appearing in the Disease History section do not share the same context as those appearing in the Discharge Treatment section.

The version we present here is not yet complete. As we have already mentioned, our final goal is to model clinical events in psychiatry. Hence the need to include a temporal representation of medical knowledge [3].

2.3. Semantic Annotation

Several tools have been developed in the field of natural language processing and semantic annotation that can be used for French texts. Among them are GATE [4], ECMT (*http://ecmt.chu-rouen.fr*) and SIFR [2] annotator. In our proposed approach, we use GATE which provides different components for semantic information extraction. We used components previously developed by the OnBaSsam [5] project and subsequently improved to meet the needs of psychiatry-related texts. Our pipeline is composed of

[3] A medical classification list by the WHO https://icd.who.int/browse10/2019/en

[4] This branch of the ontology is available at the following address https://github.com/AouinaOns/PartOfDocuments_ontology

several Processing Resources (PRs) that run sequentially on a given document, as shown in Figure 1.

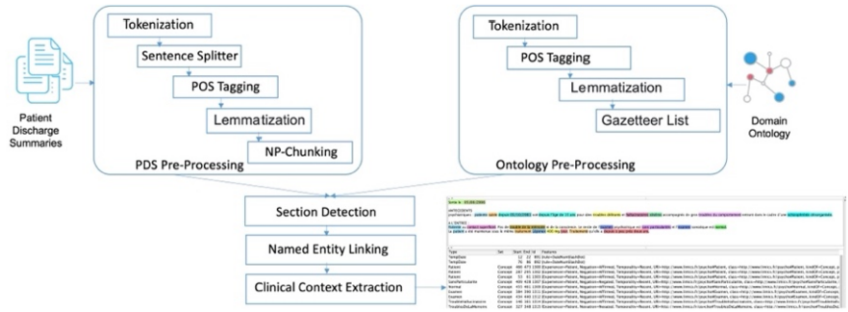


Figure 1. Illustration of the ontology-based semantic annotation process.

Task 1 - Pre-processing: First, Word tokenization is applied on PDSs, then sentence splitting, followed by Part Of Speech (POS) tagging and lemmatization, and finally, Noun Phrases extraction or NP-chunking.

For this purpose, we used Gate Corpus Pipeline, a Gate component for corpus processing, to which we added the following PRs: a French tokenizer for tokenization and the french TreeTagger[5] for annotating texts with information about parts of speech and lemmas. It has been successfully used to tag French texts. Finally, to retrieve noun phrases from PDSs, OpenNLP[6] is used, and adapted to the French language. This step allows us to build a list of noun phrases from the output of the TreeTagger. In our solution, we assume that entities such as signs and symptoms, diseases, disorders, and clinical events are noun phrases [5]. Regarding the ontology processing, we follow the same steps for the building of the gazetteer list. This list consists of the ontology concepts and their labels, pre-processed according to the steps presented in figure 1, as well as the URIs.

Task 2 - Section detection: The discourse structure of a document can be very useful for improving information retrieval tools. The identification of sections, e.g., Current Illness Story, Family Story, is crucial in our context because it is the major key to identifying the temporal context of narrative passages. As mentioned earlier, the PDSs are organized according to taxonomy (Cf. sec. 2.2). This structure is not always followed, some sections may be missing, merged or in a different order, and the same section may have several titles. For example, for the section "*family history*" we may find "*fam history*", "*psychiatric family history*", etc.

Therefore, we apply JAPE [7] rules and fuzzy string matching on the terms of the previously constructed section name gazetteer to identify section boundaries. It uses the concept of the Levenshtein distance which is a metric representing the number of character changes between words. This rule identifies the beginning of a section by the appearance of a term available in the terminology and the end before the beginning of the next term.

Task 3 - Named Entity Extraction and Linking: For this task, we classify the NP-chunk candidates obtained in the pre-processing phase to ontology concepts by assigning

[5] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger

[6] https://github.com/GateNLP/gateplugin-OpenNLP

them a URI. For this task, we develop a rule-based system combined with a fuzzy matching string classification. We define the annotation acceptance threshold as being proportional to the length of the dictionary term.
For extracting temporal entities, we used the TIMEX[7] GATE plugin to annotate the documents with TIMEX3 tags using the SUTime (Stanford Temporal Tagger) library. This allowed us to extract dates, duration e.g. "*since about 1 month*", and frequencies e.g. "*2 times a day*".

Task 4 - Clinical Context Extraction: In addition to extracting the named entities themselves, it is necessary to identify the context in which they appear in the text. Documents are pre-annotated and mapped to ontology concepts (signs, diseases, treatments, symptoms, temporal expressions, etc.) using the process described above. Then, the French FastContext [7] algorithm is applied, to identify the context of the clinical conditions annotated in a sentence. It considers three contexts: negation, hypothesis, and determination of the subject, whether the patient, patient's relative or healthcare professional.

3. Results

Table 1. Quantitative results of named entity extraction evaluations by the 2 annotators.

		Quantity	Precision	Recall	F1
Sign Or Symptom		1747	0.9544	0.9503	0.9524
Disease		150	0.9826	0.7635	0.8593
Trouble		459	0.9894	0.9493	0.9690
Clinical Event.		1459	0.9744	0.8140	0.8870
Personal Situation		188	0.9895	0.8468	0.9126
Drug	Drug Name	1034	0.8200	0.9805	0.8822
	Drug Dose	650	0.9848	0.9610	0.9727
Temporal Inf.	Date	840	0.9942	0.9709	0.9824
	Duration	529	0.9574	0.9777	0.9674
	Frequency	212	0.9459	0.8373	0.8883

The evaluation of our approach consists in manually analyzing the named entity extraction phase. Our evaluation corpus consists in 50 PDSs randomly extracted from the dataset (4100 sentences, 4213 non-unique ontology concepts annotated). Two persons evaluate the annotations and their context including negation, hypothetical, temporality, and experiencer. We have, therefore, grouped them into 10 unique higher-level concept ontologies to facilitate manual evaluation.

The precision, recall, and F-measure are presented in Table 1. An inter-annotator agreement has been calculated (0.88). The results were promising with an overall precision value of 0.9674, a recall of 0.9780, and a F1 of 0.9727.

4. Discussion and Conclusion

The objective of this work is to reconstruct structured patient data from PDSs to complete

[7] https://github.com/pkourdis/gateplugin-SUTime

the patient data in the PsyCARE project. In this paper we presented a first step which consists in semantically annotating the French psychiatry PDSs. To this end, from an unstructured set, we were able to perform a semantic annotation using Gate plugins and algorithms that we modified to fit PDSs structure.

A first evaluation of the semantic annotations shows that we are able to correctly identify the concepts of the ontology with GATE and modified algorithms to take into account French texts. These good results can be explained in particular by taking into account the structure of the PDSs, which could prove to be a weakness for other types of psychiatric documents.

The next steps in our work are to integrate into the ontology the clinical events and non-psychiatric diseases most frequently present in the PDSs. Then, it will be necessary to identify the relationships between the concepts. Given the state of the art, this will be done with machine learning and deep learning algorithms.

Acknowledgment

This work has been supported by the French government's “Investissements d'Avenir” program, which is managed by the Agence Nationale de la Recherche (ANR), under the reference PsyCARE ANR-18-RHUS-0014.

References

[1] Souaiby L, Gaillard R, Krebs MO. Durée de psychose non traitée: état des lieux et analyse critique [Duration of untreated psychosis: A state-of-the-art review and critical analysis]. 2016 Aug;42(4):361-6.
[2] Tchechmedjiev A, Abdaoui A, Emonet V, Zevio S, Jonquet C. SIFR annotator: ontology-based semantic annotation of French biomedical text and clinical notes. BMC Bioinformatics. 2018.
[3] Hilbey J, Aimé X, Charlet J. Temporal Medical Knowledge Representation Using Ontologies. Challenges of Trustable AI and Added-Value on Health: Proceedings of MIE 2022. 2022 May 1;294:337.
[4] Cunningham H, Maynard D, Bontcheva K, Tablan V, Aswani N et al. 2011. Text processing with GATE (Version 6). Sheffield: GATE. doi: https://doi.org/10.1016/j.procs.2014.05.444.
[5] Cardoso S, Meneton P, Aimé X, Meininger V, Grabli D, et al. Use of a modular ontology and a semantic annotation tool to describe the care pathway of patients with amyotrophic lateral sclerosis in a coordination network. PLOS ONE 16(1): e0244604. doi: https://doi.org/10.1371/journal.pone.0244604
[6] Zhang S, Elhadad N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. Journal of biomedical informatics. 2013 Dec 1;46(6):1088-98.
[7] Yacoubi Ayadi N, Charrad M, Vidal ME, Ben Ahmed M, Amdouni S. A Semantic Framework for Web service Annotation, Matching and Classification in Bioinformatics. Information interaction intelligence. 2011 Nov 1;11(2).
[8] Mirzapour M, Abdaoui A, Tchechmedjiev A, Digan W, Bringay S, Jonquet C. French FastContext: A publicly accessible system for detecting negation, temporality and experiencer in French clinical notes. Journal of Biomedical Informatics. 2021 May 1;117:103733..

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230269

The Prevalence of mRNA Related Discussions During the Post-COVID-19 Era

Dimitrios KOKKINAKIS[a,1], Bastiaan BRUINSMA[b] and Mia-Marie HAMMARLIN[c]
[a]*University of Gothenburg and the Centre for Ageing & Health, AgeCap, Sweden.*
[b]*Chalmers University of Technology, Sweden*
[c]*University of Lund, Department of Communication and Media, Sweden*
ORCiD ID: Dimitrios Kokkinakis https://orcid.org/0000-0003-3579-4739

Abstract. Vaccinations are one of the most significant interventions to public health, but vaccine hesitancy and skepticism are raising serious concerns for a portion of the population in many countries, including Sweden. In this study, we use Swedish social media data and structural topic modeling to automatically identify mRNA-vaccine related discussion themes and gain deeper insights into how people's refusal or acceptance of the mRNA technology affects vaccine uptake. Our point of departure is a scientific study published in February 2022, which seems to once again sparked further suspicion and concern and highlight the necessity to focus on issues about the nature and trustworthiness in vaccine safety. Structural topic modelling is a statistical method that facilitates the study of topic prevalence, temporal topic evolution, and topic correlation automatically. Using such a method, our research goal is to identify the current understanding of the mechanisms on how the public perceives the mRNA vaccine in the light of new experimental findings.

Keywords. vaccine hesitancy, structural topic modeling, Swedish internet forum, natural language processing, Swedish tweets, mRNA vaccines, event detection

1. Introduction

Vaccine hesitancy and skepticism can be triggered by anxiety about possible side effects and concerns related to novel vaccine technologies, such as the messenger RNA (mRNA) which can be used as a reason for not receiving (the COVID-19) vaccine [1]. The University of Lund study: "Intracellular Reverse Transcription of Pfizer BioNTech COVID-19 mRNA Vaccine BNT162b2 In Vitro in Human Liver Cell Line" [2], published on the 25th of February 2022, has been frequently cited since its release, as a confirmation for the reason for vaccine hesitancy, highlighting a potential misconception that the mRNA vaccine alters the human DNA. Furthermore, vaccine skepticism is often taken online on social networking sites. Therefore, our study aims to identify and discuss Swedish mRNA-related social media posts, that are being published during the period after the Lund study. We use structural topic modeling (STM) to explore Swedish tweets and Swedish discussion posts from a popular social media platform (Flashback) about mRNA-related vaccination. Our aim is to give answer to two major questions: *what patterns emerge in Swedish social media as a response to the Lund study* and *how mRNA*

[1] Corresponding Author: Dimitrios Kokkinakis, Department of Swedish, Multilingualism, Language Technology. Box 200, 405 30 Gothenburg, Sweden. E-mail: dimitrios.kokkinakis@gu.se.

related discussion topics change over time, to better understand the public perceptions, responses and concerns that arise after the Lund research. We start by applying natural language processing methods to pre-process and classify the mRNA-related narratives and then structural topic modeling to uncover the most prevalent discussion topics and their evolution over time in the Swedish context. STM automatically detects latent topics in the dataset which can be used to investigate the nature of these topics reflected in the mRNA discussions [3] by utilizing an exploratory mixed quantitative-qualitative approach using data collected from Swedish social media. As a methodological approach, STM enable us to identify prevalent topics in the data, followed by a qualitative analysis on the most representative words and posts of each topic which provided us with better and more targeted insights into the pros and cons of public perceptions and concerns about the mRNA vaccine. Our results could be useful to public health experts and pro-vaccine organizations to formulate even more effective policies and strategies to reduce anti-vaccine reactions and boost vaccine acceptance.

2. Vaccine Hesitancy and Skepticism

According to the World Health Organization [4] vaccine hesitancy, "the reluctance or refusal to vaccinate despite the availability of vaccines", was one of the top ten threats to global health even before the COVID-19 pandemic. Although vaccinations are considered as one of the most significant interventions to public health, vaccine hesitancy and resistance creates serious concerns for a significant portion of the population in many countries, including Sweden. Skepticism about vaccine effectiveness, adverse effects, personal beliefs and conspiratory claims as well as exposure to misinformation, plays an important role in decreasing rates of vaccination [5,6]. Vaccine hesitancy discussions are often taken online and for an increasing number of people, the use of such platforms has become a major source for information related to health protection and vaccinations [7]. The availability of massive digital content in e.g., Twitter or Reddit enables researchers to rapidly analyze and monitor large amounts of data, to e.g., identify and better understand the vaccine-deniers' arguments against vaccinations which in turn, can rapidly be spread as *rumours* to an even wider audience. False and disputed news or misleading information about vaccination keeps emerging and flowing between people in social media [8]. There is a sense of freedom of self-expression in the use of language, indicated by e.g., the magnitude of (negative) ways to refer to vaccine; here are examples from the Swedish data: *fejkvaccin* 'fake vaccine', *bluffvaccin* 'hoax vaccine', *fuskvaccin* 'fraud vaccine', *förtryckarvaccin* 'oppressor vaccine' or *försökskaninvaccin* 'guinea pig vaccine'.

3. Data

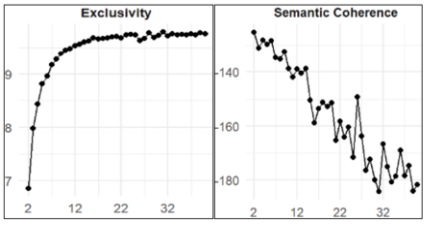
Swedish tweets were downloaded from Feb., 10, 2022 (two weeks before the Lund study was published) to Nov., 10, 2022. The tweets were collected with the keywords *m-?RNA.** ('?' the preceding character is optional.; '.*': ≥ 0 characters) or the hashtag *#mRNA* and *lang:sv* (Swedish content). The final tweet data set consisted of 1,700 unique tweets from 730 different users. Apart from the previous, we also collected ca 7,600 unique posts from the popular Swedish forum Flashback (https://www.flashback.org/), from 18 different discussion threads, all related to COVID-19 and mRNA vaccination.

3.1. Preprocessing and data cleansing

We preprocessed the data using R 4.2.1. For each tweet and for each post, we stored the text and some relevant metadata such as the date of publication. Duplicate data as well as Swedish stop words and numbers were deleted, while the textual content was turned to lowercase. During a normalization process, identified token variants such as 'mrna vaccin' and 'mrnavaccin' were converted to a single uniform format, here 'mrna-vaccin'. Furthermore, the dataset was further tokenized (basically separating punctuation and metadata from words). Multiword expressions, statistically significant collocations and phrasal verbs were also recognized, and their contiguous components were joined with an underscore prior to further processing (e.g., *big_pharma*; *in_vitro* or *spruta_in* 'to inject'). Posts with less than 3 tokens were removed due to a small search volume. For the structural topic modeling we used the R package stm (version 1.3.6) [9].

4. Structural Topic Modeling

Latent Dirichlet Allocation (LDA) [10], is a popular topic modeling method that uses the statistical analysis of textual data to identify themes or topics that occur in a document collection. Structural topic model (STM) has emerged as an extension to LDA allowing the integration of covariates into the prior distributions for document-topic compositions and topic-word proportions. Thereby, STM can be used to model how the content of a collection of documents changes as a function of document-level covariates such as day and time, and gain insights and understanding on how topics evolve. Since there is no "correct" solution for determining the optimal number of topics *k* that should be generated during the model selection process, several diagnostic aspects of the topic modeling were evaluated to decide the number of topics, *k*, to use. The *stm* package implements several evaluation metrics, such as the spread of *semantic coherence* [11] and *exclusivity*, which both capture what humans qualitatively perceive as good topics [9]. After preprocessing of the data, a document-term matrix was created with 7,600 documents, 18,900 terms (i.e., unique words) and used for modeling, while the best model yielded 14 topics. Figure 1a shows the semantic coherence vs the exclusivity of the models, while figure 1b shows the temporal evolution of two of the identified topics.



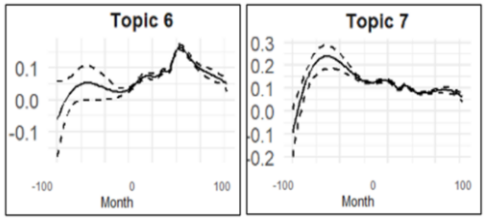


Figure 1a. Semantic coherence and selectivity; **figure 1b.** Prevalence of topics 6 and 7.

5. Results and Discussion

To summarize and better understand the public responses and concerns that arose after the Lund study, we took a closer look at the results of the STM, in which several general

themes were revealed based on the topics identified. With respect to our first research question, that is *what patterns emerge in Swedish social media as a response to the Lund study*, the most prevalent theme during the first couple of months after its publication, was, as expected, direct related to the Lund article:

- *Oroväckande resultaten från svenska studien: Pfizer-vaccin tar sig in i leverceller - och omvandlas till DNA* 'Alarming results from the Swedish study: Pfizer vaccine enters livercells - and is converted into DNA'.

A major concern was also the future unknown effects of the mRNA vaccine:

- *Risken är stor att barnen får svåra skador som kan påverka framtida generationer.* 'The risk is high that children will suffer serious injuries that could affect future generations.'.

Related to the previous is the evidence of the conformity with people's willingness to get vaccinated, but not with the mRNA vaccine, but rather with a "conventional" one:

- *Jag är absolut ingen vaccinmotståndare, så länge de är traditionellt tillverkade på ett avdödat virus. Men dessa mRNA vacciner är jag väldigt skeptisk till. Tar dom helt enkelt inte!* 'I am absolutely not against vaccines, as long as they are conventional made from a dead virus. But I am very skeptical about these mRNA vaccines. Just don't take them!'.

Concerns were raised for possible injuries and/or side effects, e.g., on the male genitalia; on the female ovaries and for the traces of mRNA vaccine in breast milk – these were major themes for the period of August-November 2022, due to new research studies:

- *Biverkningar av mRNA-vaccin inkluderar allvarliga skador på penis* 'Side effects of mRNA vaccines include serious damage to the penis'.

Finally, myocarditis risks were prominent during almost the whole examined period:

- *Läkare varnar för att mRNA vaccin orsakar myokardit men Twitter stämplar inlägget som falsk information.* 'Doctors warn that mRNA vaccine causes myocarditis but Twitter labels the post as false information.'.

Table 1 shows examples of two of the most prevalent topics with their top-weighted word content as well as an example of a "document"/post, for these topics, which address: the mRNA-to-DNA study; and mRNA's concerns and risks.

Table 1. Examples of top-rated STM topics

Topic#	Theme	Top words	Top-weighted documents
7	mRNA-to-DNA study	*spikeprotein, dna, studie, cell, levercell, omvandla, producera, virus, ta_sig_in*	*Och vem betalar till människor som fått skador 😡 Pfizer Vaccine Becomes DNA in Liver Cells.* 'And who pays for people who suffered injuries 😡 Pfizer Vaccine Becomes DNA in Liver Cells.' *frågan är OM några av oss kommer att kunna återställa det nu när 80% av jordens befolkning även har accepterat förändring av sitt DNA* 'the question is IF some of us will be able to restore it now that 80% of the earth's population has also accepted changing their DNA'
6	Concerns and risks	*myokardit, män, hjärtmuskelinfl., novavax, skada, penis, antivax, studie, foliehatt*	*Covid-19-vaccinets mRNA i bröstmjölk* 'The covid-19 vaccine's mRNA in breast milk' *Ja vaccinet samlas i äggstockarna så alla kvinnor bör undvika sex i minst 10 år.* 'Ya the vaccine accumulates in the ovaries so all women should avoid sex for at least 10 years.'

With respect on *how discussion topics change over time* some topics showed some clear characteristics on the topic prevalence (cf. Fig 1b). The time covariate provides a means for a direct comparison and explanatory power over time which makes it better understand people's concerns during a time span. Retrospectively, the variation of topics across time, can also be helpful to detect significant events related to e.g., mRNA discussions in social media data over time, as in our case, e.g., through an examination of prominent words within each topic. For instance, figure 1b shows how the prevalence of topic 6 raises towards the end of the examined period, while topic 7 rapidly declines

during the middle of the examined period; this is most probably due to the fact that topic 6 relates to the discussion of the potential side effects and risks (e.g., traces of mRNA in breast milk); while topic 7 directly relates or refers to the University of Lund study.

6. Conclusions

This study identifies dominant topics about mRNA vaccine–related issues discussed on Swedish social media. It also examines the changes in these topics over time to better understand the larger trends. Among the selected topics, certain themes, e.g., on myocarditis, remained constant over time. As vaccine development progressed however, other topics became more dominant, driven by e.g., scientific studies introduced to the public. A limitation of the presented work is the search itself which only used a non-exhaustive list of keywords; several relevant posts are probably not included. Consequently, we could expect discussions to be different when using a different keyword set. As a future task, we plan to integrate sentiment analysis. Subsequently, we could use polarity as a covariate to also capture how the sentiment of the mRNA-related events evolve over time. Although it is not the focus of this study, a closer examination of the users *might* also provide some meaningful information on whether certain users are more likely to post or comment on certain topics (cf. [12]).

Acknowledgements

This work is part of the project Rumour Mining (MXM19-1161:1) financed by the Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond).

References

[1] Leong C, Jin L, Kim D, et al. Assessing the impact of novelty and conformity on hesitancy towards COVID-19 vaccines using mRNA technology. Com Med. 2022, 2:61. doi:10.1038/s43856-022-00123-6
[2] Aldén M, Olofsson Falla F, et al. Intracellular Reverse Transcription of Pfizer BioNTech COVID-19 MRNA Vaccine In Vitro in Human Liver Cell Line. Mol. Bio. 2022, 44, 1115–1126.
[3] Scannell D, Desens L, Guadagno M, et al. COVID-19 Vaccine Discourse on Twitter. J Health Com. 2021, 26:7, 443-459. doi: 10.1080/10810730.2021.1955050
[4] World Health Organization. Ten threats to global health in 2019, 2019. [cited 2023 Jan 03]. Available from: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019
[5] Rutjens BT, van der Linden S, et al. Science skepticism in times of COVID-19. Group Processes & Intergroup Relations. 2021. 24(2), 276–283. doi:10.1177/1368430220981415
[6] Garett R, Young SD. Online misinformation and vaccine hesitancy. Translational Behavioral Medicine. 2021. 11(12):2194-2199. doi:10.1093/tbm/ibab128
[7] Klimiuk K, et al. Vaccine misinf. on social media–topic-based content & sentiment analysis of Polish vaccine-deniers' comments on FB. Hum Vacc&Imm. 2021. 17:7. doi:10.1080/21645515.2020.1850072
[8] Ahmed W, Vidal-Alaball J, et al. COVID-19 and the 5G conspiracy theory: Social network analysis of twitter data. J Med. Int Res. 2020. 22, e19458. doi:10.2196/19458
[9] Roberts M, Brandon S, Tingley D. Stm: An R package for structural topic models. J of Statistical Software. 2019 91.1: 1–40. doi:10.18637/jss.v091.i02
[10] Blei DM, et al. Latent Dirichlet Allocation. J Mach Learn Res. 2003 3:993–1022. doi:jmlr.2003.3.4-5.993
[11] Mimno D, Wallach HM, et al. Optimizing Semantic Coherence in Topic Models. Empirical Methods in NLP; 2011 July 27–31; Edinburgh, UK, c Association for Comp. Ling., p. 262–272.
[12] Lyu JC, Le Han E, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. J Med Inter Res. 2021 Jun 29;23(6):e24435. doi:10.2196/24435.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230270

Classification of Clinical Notes from a Heart Failure Telehealth Network

Fabian WIESMÜLLER[a,b,c,1], Aaron LAUSCHENSKI[b], Martin BAUMGARTNER[b,c], Dieter HAYN[a,b], Karl KREINER[b], Bettina FETZ[d], Luca BRUNELLI[e], Gerhard PÖLZL[e], Bernhard PFEIFER[f,g], Sabrina NEURURER[f,g] and Günter SCHREIER[b,c]

[a] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[b] *AIT Austrian Institute of Technology, Graz, Austria*
[c] *Institute of Neural Engineering, Graz University of Technology, Graz, Austria*
[d] *Landesinstitut für Integrierte Versorgung – LIV Tirol, Innsbruck, Austria*
[e] *Department of Internal Medicine III, Cardiology and Angiology, Medical University Innsbruck, Innsbruck, Austria*
[f] *Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria*
[g] *Division for Digital Medicine and Telehealth, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall (Tyrol), Austria*
ORCiD ID: Wiesmüller Fabian https://orcid.org/0000-0001-6567-7782

Abstract. Heart failure is a common chronic disease which is associated with high re-hospitalization and mortality rates. Within the telemedicine-assisted transitional care disease management program HerzMobil, monitoring data such as daily measured vital parameters and various other heart failure related data are collected in a structured way. Additionally, involved healthcare professionals communicate with one another via the system using free-text clinical notes. Since manual annotation of such notes is too time-consuming for routine care applications, an automated analysis process is needed. In the present study, we established a ground truth classification of 636 randomly selected clinical notes from HerzMobil based on annotations of 9 experts with different professional background (2 physicians, 4 nurses, and 3 engineers). We analyzed the influence of the professional background on the inter annotator reliability and compared the results with the accuracy of an automated classification algorithm. We found significant differences depending on the profession and on the category. These results indicate that different professional backgrounds should be considered when selecting annotators in such scenarios.

Keywords. Clinical notes, Annotation, Text classification, Natural Language Processing

1. Introduction

Due to the aging population the need for adequate treatment of chronic diseases becomes an ever more pressing issue. One of the most severe chronic illnesses is heart failure,

[1] Corresponding Author: Fabian Wiesmüller, AIT Austrian Institute of Technology, Graz, Austria, E-Mail: fabian.wiesmueller@ait.ac.at.

with a 12-month all-cause hospitalization rate of 44% for hospitalized and 32% for ambulatory heart failure patients [1,2]. To improve the outpatient care of heart failure patients and to reduce the hospitalization rates the telemedical disease management program HerzMobil was developed by the AIT Austrian Institute of Technology in cooperation with Landesinstitut für Integrierte Versorgung – LIV Tirol, the UMIT TIROL - Private University for Health Sciences and Health Technology, Hall (Tyrol), the telbiomed Medizintechnik und IT Service GmbH and the Tirol Kliniken [2]. During a three-month period after hospital discharge the patients track selected vital parameters like e.g., the blood pressure and upload them on a daily basis. Whilst most of the data are transmitted in a structured form, healthcare professionals (HCP) can additionally communicate via free text clinical notes. These notes can contain valuable information, such as reasons for medication changes, the absence of a patient or an HCP or a contact with a patient [2]. However, this information is currently only available in an unstructured format and thus impractical for analysis. Due to the rapidly increasing number of clinical free texts, various groups made efforts to work towards an automated analysis of such texts by analyzing and classifying clinical free texts by natural language processing (NLP), a subfield of machine learning [3-8]. In previous studies, automatic extraction of date and time references [9] and classification into predefined categories [10] were performed on notes of the HerzMobil program. These NLP solutions aim to improve the workflow of HerzMobil by providing additional information and reducing manual work (e.g., with filter functions).

To develop and train supervised machine learning models, a set of annotated data with an established ground truth is necessary. Therefore, a subset of the available HerzMobil notes had to be manually annotated and classified into categories.

This work focuses on the annotation process of clinical notes and evaluates the inter-observer agreement in between different professional groups (physicians, nurses, engineers). Additionally, the annotations are compared with the classifications of a regular expression-based machine learning model developed for this work.

2. Methods

2.1. Dataset

All the notes used for this work have been de-identified and split into their individual sentences before any further processing, based on a pre-existing algorithm [11]. This resulted in 636 individual text snippets for the annotation process. In the following the word note will refer to such a text snippet on a sentence level.

2.2. Note categories

In a multi-step, user-centered process with various HerzMobil stakeholders, eight different categories were identified to be of high interest for the involved personnel: *Absence, Home visitation, Contact HCP, Contact patient, Contact others, Education, Technical problems,* and *Therapeutic regime*. Each note could be assigned to zero, one or multiple categories.

2.2.1. Annotators and Guideline

Nine professionals from the project team, all familiar with the HerzMobil system, annotated the entire dataset: two physicians, four telehealth nurses, and three engineers. An annotation guideline was developed containing definition of the category. Throughout the user-centered process of developing the categories, the annotation guide was updated multiple times to reduce vagueness and result in clear category descriptions. The guide has been applied by the experts in three iteration steps: During the first iteration eight of the nine annotators annotated all 636 notes by either ticking or not each category for each note, which resulted in 0 to eight ticks per note. After this iteration, based on those notes and categories that featured inhomogeneous classification results, the annotation guide was adapted to clarify discrepancies and to reduce ambiguity. For the annotation of the notes, a pre-existing annotation tool was used [12].

With this updated guide, a second round of annotations was conducted by two engineers independently, one from the previous group of eight experts and one additional scientist (annotator number nine). During the second iteration, not only the notes themselves but also the ticks from the first iteration were considered by the annotators, while applying the new version of the guideline.

In a third iteration, those notes that were annotated differently by the two engineers were manually inspected, whereas the third engineer of the annotator team decided whether a category should be ticked or not to establish the final ground truth.

2.3. Statistical Analysis

The inter annotator reliability was calculated through the Cohen's Kappa (κ) by comparing the annotations of each annotator with every other annotator. The κ was used since it is one of the standard tools to measure reliability on binary classification tasks [13]. The κ was calculated per category as well as per annotator role, to evaluate whether specific categories were annotated with less ambiguity and if a difference in the annotator roles was noticeable. Subsequently, the agreement of the individual annotators with the ground truth was analyzed. Additionally, an automated classification algorithm, based on regular expressions, was compared to the annotators. The level of significance was determined with the Friedman test. $p < 0.05$ was considered statistically significant. The statistical analyses were performed with the Predictive Analytics Toolbox for Healthcare [14] based on MATLAB (The MathWorks, Nattick, MA). The studies were approved by the ethics committee of the Medical University Innsbruck (vote nr. 1035/2022).

3. Results

3.1. Comparing Roles

Table 1 shows the reliability between the three roles over all eight categories. A highly significant difference ($p < 0.0001$) between the professions has been identified.

Table 1. Comparison of the Cohen's Kappa between the different roles, combined over all categories.

	Engineers	Nurses	Physicians
Engineers	0.606	0.447	0.386
Nurses	0.447	0.371	0.352
Physicians	0.386	0.352	0.270

3.2. Comparing categories

The analysis of the annotations per category is shown in Table 2 and resulted in a highly significant ($p < 0.0001$) difference between the categories using the Friedman test.

Table 2. Comparison of the Cohen's Kappa between the different categories over all eight annotators.

Metrics	Absence	Home visitation	Contact HCP	Contact patient	Contact others	Training	Technical problems	Therapy regime
Median	0.288	0.557	0.321	0.433	0.282	0.335	0.553	0.527

3.3. Comparison with the ground truth

The comparison of a) each annotator and b) the regular-expression-based algorithm with the established ground truth is displayed in Table 3. Cohen's Kappa differed significantly between categories ($p < 0.01$), whilst no significant difference has been proven when aggregated per annotator ($p = 0.107$).

Table 3. Comparison of the individual annotators and the regular expressions with the ground truth, using the Cohen's Kappa (κ). Eng... Engineer, Phy... Physician, Nur... Nurse, ReEx... Regular Expressions

Category	Eng 1	Eng 2	Phy 1	Phy2	Nur 1	Nur 2	Nur 3	Nur 4	Total *	RegEx
Absence	0.858	0.860	0.000	0.484	0.573	0.556	0.405	0.085	0.520	0.720
Home visitation	0.872	0.969	0.381	0.708	0.773	0.357	0.844	0.656	0.741	0.912
Contact others	0.536	0.512	0.191	0.119	0.171	0.589	0.502	0.533	0.507	0.158
Contac HCP	0.714	0.400	0.329	0.629	0.485	0.690	0.131	0.667	0.557	0.147
Contact patient	0.277	0.182	0.106	0.167	0.323	0.241	0.128	0.425	0.212	0.050
Education	0.353	0.674	0.688	0.320	0.648	0.743	0.514	0.042	0.581	0.191
Technical problems	0.496	0.400	0.529	0.753	0.533	0.538	0.736	0.533	0.533	0.824
Therapy regime	0.511	0.640	0.413	0.359	0.617	0.659	0.491	0.659	0.564	0.048
Total**	0.524	0.576	0.355	0.422	0.553	0.572	0.496	0.533		0.174

* Median κ of all annotators per category ** Median κ of all categories per annotator

4. Discussion and Outlook

For this work free text clinical notes from a heart failure telehealth program have been annotated by nine experts using an annotation guide. These annotations have been analyzed regarding the agreement between the annotators among themselves and between the annotators and the ground truth.

Table 1 shows that the engineers had the highest agreement out of the three expert roles over all categories, whilst the least coherent annotations were done by the physicians with a difference in the median of κ of 0.336. It can also be seen that comparing the physicians and nurses with the engineers, results in a higher reliability than the respective groups achieved themselves. This result must be put into perspective, since even though the nurses and physicians contributed significantly to the annotation guide, it was mainly developed by the engineers, which ultimately means that the engineers invested more time in the annotation process.

Table 2 shows that the reliability of the annotators varies throughout the categories which was also indicated by a highly significant difference yielded by the Friedman test. Table 2 also shows that the reliability of commentators varied across categories, as

confirmed by a highly significant Friedman test. This confirms our assumption that some categories are easier to define and categorize than others. For example, *Home visitation*, *Technical problems* and *Therapeutic regime* were all classified with a high reliability ($\kappa > 0.525$) whilst *Absence* and *Contact with others* were annotated with less agreement ($\kappa < 0.290$). This can be explained by the nature of the categories, since *Technical problems* had a smaller margin of interpretation and could therefore be defined rather precisely compared to categories like *Education* which includes a broader range of topics.

The comparison with the ground truth shown in Table 3 revealed that the regular expressions were able to outperform the annotators drastically in the categories *Absence, Home visitation* and *Technical problems* with a respective difference in κ of 0.200, 0.171 and 0.291. However, the algorithm did not surpass the manual annotations in the remaining five categories. This prototypical regular expression algorithm will, however, be further developed and optimized on the HerzMobil notes in future work.

The knowledge gained in this work provides a helpful basis for future annotation tasks. For example, the established annotation guide can be used to annotate larger data sets that can be used as regular expressions in future projects for training and validating machine learning models. This would benefit not only the workflow in the HerzMobil program, but also other telehealth systems with a similar accumulation of clinical notes.

Acknowledgements: Parts of this work were supported by the Land Tirol, in the framework of the project “d4Health Tirol”.

References

[1] Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2016 Jul;37(27):2129-2200.
[2] Von der Heidt A, Ammenwerth E, Bauer K, et al. HerzMobil Tirol network: rationale for and design of a collaborative heart failure disease management program in Austria. Wien Klin Wochenschr. 2014 Nov;126(21-22):734-741
[3] Kong HJ. Managing Unstructured Big Data in Healthcare System. Healthc Inform Res. 2019 Jan;25(1):1-2.
[4] Mora S, Attene J, Gazzarata R, et al. A NLP Pipeline for the Automatic Extraction of Microorganisms Names from Microbiological Notes. Stud Health Technol Inform. 2021 Oct 27;285:153-158
[5] Yehia E, Boshnak H, AbdelGaber S, et al. Ontology-based clinical information extraction from physician’s free-text notes. J Biomed Inform. 2019;98:103276.
[6] Lee HJ, Zhang Y, Jiang M, et al. Identifying direct temporal relations between time and events from clinical notes. BMC Med Inform Decis Mak. 2018;18(Suppl 2):49.
[7] Fernandes MB, Valizadeh N, Alabsi HS, et al. Classification of neurologic outcomes from medical notes using natural language processing. Expert Syst Appl. 2023 Mar 15;214:119171
[8] George A, Johnson D, Carenini G, et al. Applications of Aspect-based Sentiment Analysis on Psychiatric Clinical Notes to Study Suicide in Youth. AMIA Jt Summits Transl Sci Proc. 2021 May 17;2021:229-237.
[9] Wiesmueller F, Eggerth A, Kreiner K, et al. Automated Extraction of Time References from Clinical Notes in a Heart Failure Telehealth Network. Computing in Cardiology Conference (CinC). 2020.
[10] Wiesmüller F, Kreiner K, Eggerth A, et al. Natural Language Processing for Free-Text Classification in Telehealth Services: Differences Between Diabetes and Heart Failure Applications. In dHealth 2021: From eHealth to dHealth 2021 (pp. 105-109). IOS Press.
[11] Baumgartner M, Schreier G, Hayn D, et al. Impact Analysis of De-Identification in Clinical Notes Classification. Studies in Health Technology and Informatics. 2022 May 16;293:189-196.
[12] Kreiner K, Hayn D, Schreier G. Twister: A Tool for Reducing Screening Time in Systematic Literature Reviews. In: EFMI-STC; 2018.
[13] Vach W. The dependence of Cohen's kappa on the prevalence does not matter. Journal of clinical epidemiology. 2005 Jul 1;58(7):655-61.
[14] Hayn D, Veeranki S, Kropf M, et al. Predictive analytics for data driven decision support in health and care. it - Information Technology. 2018;60(4):183-94.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230271

Classifiers of Medical Eponymy in Scientific Texts

Dennis TODDENROTH[1]
Chair of Medical Informatics, University Erlangen-Nuremberg, Germany

Abstract. Many concepts in the medical literature are named after persons. Frequent ambiguities and spelling varieties, however, complicate the automatic recognition of such eponyms with natural language processing (NLP) tools. Recently developed methods include word vectors and transformer models that incorporate context information into the downstream layers of a neural network architecture. To evaluate these models for classifying medical eponymy, we label eponyms and counterexamples mentioned in a convenience sample of 1,079 Pubmed abstracts, and fit logistic regression models to the vectors from the first (vocabulary) and last (contextualized) layers of a SciBERT language model. According to the area under sensitivity-specificity curves, models based on contextualized vectors achieved a median performance of 98.0% in held-out phrases. This outperformed models based on vocabulary vectors (95.7%) by a median of 2.3 percentage points. When processing unlabeled inputs, such classifiers appeared to generalize to eponyms that did not appear among any annotations. These findings attest to the effectiveness of developing domain-specific NLP functions based on pre-trained language models, and underline the utility of context information for classifying potential eponyms.

Keywords. Entity classification, medical eponymy, transformer models.

1. Introduction

Concepts that are named after persons are a frequent characteristic of medical terminology and its textual manifestations. Such eponyms may refer to a variety of notions such as diseases, diagnostic signs, therapeutic interventions, or anatomical parts. While most medical eponyms are named after pioneering researchers (such as *Alzheimer disease*), some also originate from affected patients (*Lou Gehrig disease*) or from historical or mythological characters (*ceasarean section*). Eponyms typically resonate with scientific achievement, although their use has also been criticized for misrepresenting academic merit as well as for lacking conceptual accuracy [1].

While existing clinical and scientific texts continue to feature numerous eponyms, however, methods from computational linguistics or natural language processing (NLP) will likely encounter these peculiar phenomena. In contrast to other parts of medical terminology, eponyms are not assembled from semantic elements that provide clues to their medical meaning. Human readers who are not acquainted with a particular condition such as *Addison disease* cannot infer the involved pathophysiology from the eponym; its composed synonym *primary adrenal insufficiency*, on the other hand, is more descriptive in this regard.

[1] Corresponding Author: Dennis TODDENROTH, E-mail: dennis.toddenroth@fau.de.

Other properties like misspellings (*Fischer's exact test*) complicate the automatic detection and analysis of eponyms in conventional biomedical texts. Many clinical terms are named after several persons (*Stevens-Johnson syndrome*), while a few eminent researchers such as Harvey Cushing managed to spawn several concepts. Some terms can occur as eponyms or as non-eponyms, as in a *fractionated dosage of 50 gray* in contrast to *gray matter volume*. Near-homonyms such as *Wegner* and *Wegener* can be misused in patient records, which has motivated the development of preventative clinical decision support functions [2]. The international origin of many medical eponyms can introduce characters that are unconventional in standard English, and inconsistent Anglicization may entail variants such as *Bekhterev* and *Bechterew*. Gradual absorption into parlance may involve a loss of capitalization, as well as inflections and compositions such as *fallopian tube* and *rickettsiosis*.

Previous research has attempted to systematically collect medical eponyms, for example in order to trace their usage, including by automatically expanding search queries with variations from a curated eponym list [3]. Recent applications of advanced NLP approaches have increasingly affected knowledge synthesis from the scientific literature [4]. Such modern NLP methods include word vectors and transformer models that harness high-dimensional numeric representations derived from co-occurrence patterns trained in large textual corpora. This research studies the utility of word vectors and transformer models for recognizing medical eponymy by training and evaluating classifiers in excerpts from the medical literature.

2. Methods

Recent language models represent textual subsequences (tokens) as high-dimensional numeric vectors, or word embeddings. First-layer vocabulary vectors are thereby consistently mapped to the same locations in vector space, regardless of context. Transformer models also use positional encoding and a so-called attention mechanism to incorporate weighted information from surrounding tokens into the downstream vectors of a multilayer neural network. While vocabulary vectors of homonyms as in *Down-regulated gene functions* and *adults with Down syndrome* do not depend on preceding or subsequent tokens, contextual information embedded in the hidden layers of a transformer model thus promises to potentially improve the discrimination of ambiguous eponyms.

The following evaluation considers eponymy classifiers based on a domain-specific language model in a convenience sample of abstracts downloaded from Pubmed. Annotations were defined as either eponyms or non-eponyms using the *Brat Rapid Annotation Tool* [5]. Figure 1 illustrates the annotation procedure as well as the described context-dependent disaggregation of hidden-layer vectors. To improve the efficiency of the annotation process, abstracts were processed in five batches, and candidate labels were automatically pre-annotated with increasingly refined classifier versions. All computed candidate pre-annotations were then manually reviewed and labeled as either eponyms or non-eponyms. Initial pre-annotations were based on the cosine similarity between candidate phrases and average vocabulary vectors from an initial set of eponyms, specified as a plain list of names. Meaningful spatial relations between vocabulary vectors imply that even such a simple approach may potentially generalize to other eponyms beyond the seeded list.

Subsequent batches were pre-annotated by fitting a logistic regression model to contextualized vectors from the last hidden layer of the SciBERT language model, using the *scivocab_uncased* configuration; SciBERT has been pre-trained in a domain-specific corpus, and uses vectors with 768 dimensions [6]. Previous research suggests that models trained in a pertinent biomedical corpus may lead to better performance in domain-specific downstream tasks [7], including the disambiguation of homonyms [8]. Since the SciBERT tokenizer yields a sub-word granularity that disaggregates many eponyms (as in *kl#ats#kin*), word-related vectors were computed by averaging token vectors. Eponym annotations included entire compositions such as *Salmonella* when forming single words, while noun phrases such as *Spearman correlation* were labelled discriminately. Since the attention mechanism is limited to sequences with a maximum of 512 tokens, processing verbose abstracts required that inputs were first separated into a set of chunks, and that outputs were later re-aggregated.

The refined classifier, which potentially discriminates homonyms based on their contexts, was also applied for a plausibility check whereby annotations that appeared to conflict with in-sample predictions were manually revised. The subsequent performance evaluation considered aggregated annotations from all batches as units of observations, and calculated the area under sensitivity-specificity curves in held-out partitions of 100 bootstrapping subsamples. Scripts for computing SciBERT vectors were implemented in Python 3.8, and invoked the language model via the *transformers* package from huggingface.co [9]; regression models were trained via the *glm()* function that ships with of R 3.6. The author, who is trained as a physician, defined and revised all labels. Annotations are available at https://github.com/dtoddenroth/medicaleponyms/.

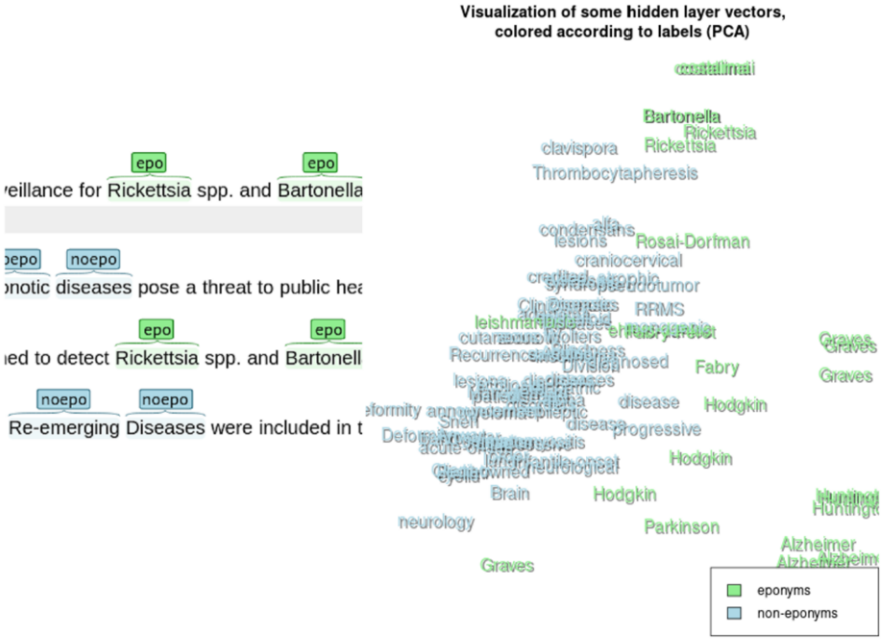


Figure 1. Screenshot of the brat tool during annotation. Candidate sequences are labeled as either eponyms or non-eponyms (left). Visualization of a principal component analysis (PCA) of hidden-layer vectors for exemplary eponyms and counterexamples. While this image shows only some of the information contained in all dimensions, we see moderate label disaggregation. Note that some eponyms appear repeated in different locations, which could allow the context-dependent disambiguation of homonyms (right).

3. Results

The described annotation procedure labelled 1,582 of 13,659 annotations in 1,079 Pubmed abstracts as eponyms (11.6%), which amounts to an average of 1.47 eponyms and 12.7 annotations per abstract. Annotated eponyms included 341 different words, while the three most frequent ones were *Fabry* (227x), *Alzheimer* (148x), and *Parkinson* (81x).

Figure 2 summarizes essential observations from the evaluation of the transformer-based eponymy models. In 100 bootstrap repetitions, logistic regression models trained on first-layer (vocabulary) vectors achieved a median area under the sensitivity-specificity curve of 95.7% (interquartile range 95.4% - 96.1%). Comparable models trained on contextualized vectors from the last hidden layer achieved a median area of 98.0% (97.7% - 98.2%), thus outperforming first-layer models by a median of 2.3 percentage points.

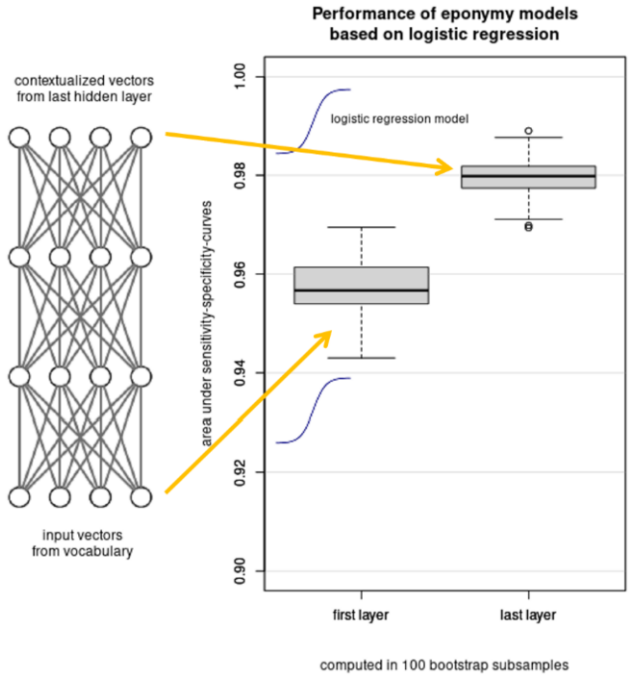


Figure 2. Distribution of the observed performance of eponymy models trained on vocabulary vectors and on contextualized hidden-layer vectors, each computed from 100 bootstrap subsamples. Note that shown y-axis is restricted to areas under sensitivity-specificity curves between 90% and 100%.

During the described batch-wise annotation procedure it also became apparent that intermediate classifier versions successfully generalized to reasonable pre-annotations that did not appear among the previous labels, including *Bland*, *Altman*, *Hounsfield*, and *Kalman*. These anecdotal observations also underline the general capability of the approach.

4. Discussion

The observed performance of the evaluated eponymy classifiers attests to the effectiveness of fitting *'minimal task-specific neural architectures'* to contextualized embeddings [6]. Such eponymy classifiers may be applied to automatically expand previous analyses of their use in literature over time [3]. Note that currently only some eponyms are represented in *Medical Subject Headings* (MeSH), which in Pubmed are assigned to indexed publications. If a more comprehensive set of eponyms such as *Bosworth fracture* could be systematically mapped to explanations like *distal fibula fracture with posterior dislocation of the proximal fragment*, this collection of translations might become useful for improving the reach or precision of pertinent queries to medical literature databases.

As a limitation, the evaluated classifier was restricted to deciding whether a given token sequence constitutes an eponym, and cannot yet properly delimit multi-word entities such as *Bland-Altman*. The superior recognition observed with the contextualized vectors, however, indicates the potential utility of incorporating information from phrases around candidate eponyms, and could also be instrumental for detecting the spans of multi-word eponyms. Previous research has considered various patterns that are typical in the medical literature, including hypernyms [10] and frequent abbreviations [11]. Since annotated datasets could be increasingly valuable for exploring alternative NLP methods in these settings, we hope that the distributed eponym labels might likewise become useful for further experiments.

References

[1] Woywodt A and Matteson E. Should eponyms be abandoned? BMJ. 2007 Sep 1;335(7617):424, doi: 10.1136/bmj.39308.342639.AD.

[2] Baskaran LN, Greco PJ, Kaelber DC. Case report medical eponyms: an applied clinical informatics opportunity. Appl Clin Inform. 2012 Sep 19;3(3):349-55. doi: 10.4338/ACI-2012-05-CR-0019.

[3] He L, Cornish TC, Kricka LJ, Vandergriff TW, Yancey K, Nguyen K, Park JY. Trends in dermatology eponyms. JAAD Int. 2022 Apr 18;7:137-143. doi: 10.1016/j.jdin.2022.03.006.

[4] Wang K and Herr I. Machine-Learning-Based Bibliometric Analysis of Pancreatic Cancer Research Over the Past 25 Years. Front Oncol. 2022. doi: 10.3389/fonc.2022.832385.

[5] Stenetorp P, Pyysalo S, Topic G, Ohta T, Ananiadou S, Tsujii J. brat: a Web-based Tool for NLP-Assisted Text Annotation. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 102–107, 2012.

[6] Beltagy I, Lo K, Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on NLP (EMNLP-IJCNLP). https://www.aclweb.org/anthology/D19-1371.

[7] Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, Kingsbury P, Liu H. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform. 2018 Nov;87:12-20. doi: 10.1016/j.jbi.2018.09.008.

[8] Toddenroth D. Evaluation of Domain-Specific Word Vectors for Biomedical Word Sense Disambiguation. Stud Health Technol Inform. 2022 May 16;292:23-27. doi: 10.3233/SHTI220314.

[9] Wolf T, Debut L, Sanh V, Chaumond J, Delangue, C.Moi A et al. Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020. https://www.aclweb.org/anthology/2020.emnlp-demos.6.

[10] Hearst MA. Automatic acquisition of hyponyms from large text corpora. Proceedings of the 14th conference on Computational linguistics. 1992. doi: 10.3115/992133.992154.

[11] Grossman Liu L, Grossman RH, Mitchell EG, Weng C, Natarajan K, Hripcsak G, Vawdrey DK. A deep database of medical abbreviations and acronyms for natural language processing. Sci Data. 2021 Jun 2;8(1):149. doi: 10.1038/s41597-021-00929-4.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230272

Communicating in Emergency Settings: BabelDr or Telephone Interpreting?

Céline DE GRAAF [a,1], Pierrette BOUILLON [a] and Hervé SPECHBACH [b]
[a] *Faculty of Translation and Interpreting, University of Geneva, Geneva, Switzerland*
[b] *Geneva University Hospitals, Department of Community Medicine, Geneva, Switzerland*

Abstract. In this paper, we present a study comparing two mediums that can be used to communicate with allophone patients: a speech-enabled phraselator (BabelDr) and telephone interpreting. To identify the satisfaction provided by these mediums and their pros and cons, we conducted a crossover experiment where doctors and standardized patients completed anamneses and filled in surveys. Our findings suggest that telephone interpreting offers better overall satisfaction, but both mediums presented advantages. Consequently, we argue BabelDr and telephone interpreting can be complementary.

Keywords. Interpreting, BabelDr, emergency, communication, satisfaction

1. Introduction

In healthcare settings, good communication is necessary as it impacts positively patients' recovery and emotional health [1]. In emergency departments, telephone interpreting is often used as it provides access to multiple languages and is often the most cost-effective and efficient way of communicating [2]. Phraselators can also enable medical communication while ensuring translation quality and data confidentiality. At the Geneva University Hospitals (HUG), one can use BabelDr [3]: a phraselator created specifically for triage that allows doctors to interact orally and patients to give answers using pictographs.

The primary aim of the present study was to assess patients' and doctors' satisfaction when using a phraselator or calling an interpreter. Our results highlight the mediums' features playing a role in patients' and doctors' satisfaction. It is, to the best of our knowledge, the first research comparing telephone interpreters to a medical phraselator.

2. Method

The experiment was conducted at the HUG in March 2022. It included sixteen participants: eight French-speaking voluntary doctors from the outpatient emergency department, and eight English-speaking standardized patients. Participants were then grouped into eight pairs, each including a doctor and a patient, who had to roleplay two anamnesis scenarios (disease – medium) following a crossover design (Table 1).

[1] Corresponding author: Céline DE GRAAF, E-mail: celine.degraaf@unige.ch

Before each scenario, patients were standardized. After each scenario, all participants had to fill in a survey. They were asked to comment on the advantages and drawbacks of the medium that was used, and to assess their agreement with sentences evaluating their satisfaction for different aspects (see first column of Table 2) on a five-point Likert scale (1, completely disagree – 5 fully agree).

Table 1. Crossover design.

Pair	Scenario 1	Scenario 2
1 to 4	Appendicitis – BabelDr	Cholecystitis – Interpreter
5 to 8	Appendicitis – Interpreter	Cholecystitis – BabelDr

3. Results

Overall satisfaction with all aspects studied was higher with telephone interpreters than with BabelDr (mean value of 3.87 ±1.05, against 3.26 ±1.26). Even though no medium obtained a perfect rating, interpreters were preferred over BabelDr for all aspects but confidentiality (Table 2).

Table 2. Results. Average of the five-point Likert scale values (AVG), and standard deviation (SD). N=16

Satisfaction aspect	BabelDr (AVG, SD)		Interpreting (AVG, SD)	
Confidentiality	4.50	±0.73	4.44	±0.81
Patient-doctor's contact	2.62	±1.15	3.62	±1.02
Problem understanding	3.12	±1.09	3.69	±1.14
Quality of communication	2.81	±1.17	3.75	±1.06

We received 88 comments on BabelDr (39 positives, 49 negatives), and 68 on phone interpreting (38 positives, 30 negatives). BabelDr's advantages are its interface, with pictographs and written translations, and that it can be used everywhere at any time. Its main drawback is that questions are limited and cannot be asked or answered freely. Phone interpreting pros are the cultural comfort it provides and the freedom of speech it offers. The cons are mainly the unavailability and the lack of trust in the translations.

4. Discussion and Conclusions

Our results show that, even though no medium has obtained a perfect rating, interpreting is considered more satisfactory than BabelDr. However, based on our results, we argue a tool such as BabelDr could compensate for the lack of interpreters for emergencies or simple cases.

References

[1] Chan Y-F, Alagappan K, Rella J, Bentley S, Soto-Greene M, Martin M. Interpreter Services in Emergency Medicine. The Journal of Emergency Medicine. 2010 Feb;38(2):133–9, doi: 10.1016/j.jemermed.2007.09.045.

[2] Squires A. Health translators and interpreters in national healthcare systems. In: Ji M, Taibi M, Crezee I.H.M, editors. Multicultural Health Translation, Interpreting and Communication. 1st ed. Routledge; 2019. p. 25-36, doi: 10.4324/9781351000390.

[3] BabelDr: spoken language translation of dialogues in the medical domain. Available from: https://babeldr.unige.ch/.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230273

Diagnosis Classification in the Emergency Room Using Natural Language Processing

Marieke M. VAN BUCHEM[a,1], Hanna H. 'T HART[a,b], Pablo J. MOSTEIRO[b], Ilse M.J. KANT[a] and Martijn P. BAUER[a]
[a] *Leiden University Medical Center, Leiden, The Netherlands*
[b] *University of Utrecht, Utrecht, The Netherlands*
ORCiD ID: Marieke van Buchem https://orcid.org/0000-0002-2917-0842

Abstract. Diagnosis classification in the emergency room (ER) is a complex task. We developed several natural language processing classification models, looking both at the full classification task of 132 diagnostic categories and at several clinically applicable samples consisting of two diagnoses that are hard to distinguish.

Keywords. Natural language processing, diagnosis classification, emergency medicine

1. Introduction

In the emergency room (ER), action by clinicians must be taken rapidly. Patients come in with various symptoms and illnesses, making it difficult for clinicians to diagnose a patient quickly and accurately. Errors in diagnosis occur more often at the ER than in the rest of the hospital [1], which can result in serious harm to the patient [2]. Natural Language Processing (NLP) has the potential to assist ER clinicians with diagnosing patients based on the clinical notes that they write while examining a patient. In the current study, we investigate the feasibility of developing a diagnosis classification model to assist ER clinicians in correctly and timely diagnosing a patient.

2. Methods

2.1. Data

We used the description of the history of present illness from the letter that is sent to the general practitioner after a patient visits the ER. The dataset consisted of 72.990 letters from unique encounters that took place between 2011 and 2021 at the Leiden University Medical Center. The number of unique diagnoses was 1997, which we categorized using the Clinical Classifications Software Refined tool. After removing encounters that could not be linked to an ICD-10 code, encounters that did not have a valid diagnosis (such as 'fever'), and categories that occurred less than 10 times, the number of encounters was 29871, with 132 unique diagnostic categories.

[1] Corresponding Author: Marieke M. van Buchem, E-mail: m.m.van_buchem@lumc.nl

2.2. Models

We trained a baseline model using TF-IDF and SVM and finetuned two pretrained BERT models (the Dutch BERTje and the Dutch medical MedRoBERTa.nl)[3].

2.3. Experiments

We conducted two experiments. In experiment 1, we trained the models on the full classification task, including all 132 diagnostic categories. In experiment 2, a clinician defined three sets of diagnoses that are difficult to distinguish from each other: heart failure versus Covid-19 (sample 1); biliary tract disease versus aortic peripheral and visceral artery aneurysms (sample 2); and acute hemorrhagic cerebrovascular disease versus meningitis (sample 3). Then, using only the encounters that included one of the diagnoses in a set, we trained the model to choose the correct diagnosis per encounter.

3. Results and Discussion

All models in experiment 2 outperformed the models in experiment 1. The BERT models outperformed the baseline model in all experiments, although the difference was small for experiment 2, sample 3. Within experiment 1, BERTje performed best, while within experiment 2, MedRoBERTa.nl performed best in all samples (see Table 1).

Table 1. Micro F1-score of the three different models in experiment 1 and experiment 2, samples 1-3.

Experiment number	Baseline	BERTje	MedRoBERTa.nl
Experiment 1	0.28	**0.35**	0.32
Experiment 2, sample 1	0.72	0.90	**0.91**
Experiment 2, sample 2	0.54	0.71	**0.79**
Experiment 2, sample 3	0.8	0.83	**0.86**

The current experiments show that looking at samples of diagnoses might be more feasible to develop for clinical practice than trying to create a classification model for all diagnostic categories. Within the next months, we will refine the samples we created with a larger group of clinicians. Furthermore, we will use LIME, an explanation algorithm, to explain the differences in output between the different models [4]. Lastly, we will develop multimodal models that also include structured data to optimize performance for both tasks.

References

[1] Hussain F, Cooper A, et alDiagnostic error in the emergency department: learning from national patient safety incident report analysis. BMC Emergency Medicine. 2019;19:1-9.
[2] Balogh EP, Miller BT. Committee on diagnostic error in health care. National Academies Press. 2015.
[3] Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019;1:4171–4186.
[4] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowlegde discovery and data mining. 2016;1135-1144.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230274

Predicting Depression Risk in Patients with Cancer Using Multimodal Data

Anne DE HOND[a,b,1], Marieke VAN BUCHEM[a,b], Claudio FANCONI[b,c], Mohana ROY[b], Douglas BLAYNEY[b], Ilse KANT[d], Ewout STEYERBERG[a] and Tina HERNANDEZ-BOUSSARD[b]
[a] *Leiden University Medical Center, Leiden, The Netherlands*
[b] *Stanford University, Stanford, CA, USA*
[c] *ETH Zürich, Zürich, Switzerland*
[d] *University Medical Center Utrecht, Utrecht, The Netherlands*

Abstract. When patients with cancer develop depression, it is often left untreated. We developed a prediction model for depression risk within the first month after starting cancer treatment using machine learning and Natural Language Processing (NLP) models. The LASSO logistic regression model based on structured data performed well, whereas the NLP model based on only clinician notes did poorly. After further validation, prediction models for depression risk could lead to earlier identification and treatment of vulnerable patients, ultimately improving cancer care and treatment adherence.

Keywords. Natural Language Processing, machine learning, oncology, depression

1. Introduction

Patients with cancer starting invasive treatment programs often develop depression that physicians struggle to recognize at an early stage [1,2]. We developed a prediction model for early identification of patients at risk for depression within the first month of chemo- or radiotherapy treatment to assist physicians and healthcare workers.

2. Methods

We included adult patients receiving cancer treatment from a comprehensive cancer center in the United States (2008-2022). Exclusion criteria were patients younger than 18 years, no clinician notes within the two weeks prior to treatment or a depression diagnosis within the year prior to treatment. Depression was defined as a depression diagnosis via ICD-9 and ICD-10 codes. Depression risk was predicted within one month after the start of cancer treatment. We included several structured data features from the patient's Electronic Health Record (EHR), like gender, age, cancer stage, history of depression, and patient email classification scores. We also included unstructured text data from clinician notes. Several machine learning (ML) models (e.g., LASSO logistic regression, random forest, gradient boosting decision trees) were compared to predict

[1] Corresponding Author: Anne de Hond, E-mail: a.a.h.de_hond@lumc.nl.

depression risk on combinations of the structured data. Three (multimodal) Natural Language Processing models (DistilBERT [3]) were developed on different combinations of the structured data and unstructured data. We split data randomly for all models in the same 2/3 train and 1/3 test set. Model performance was measured via the area under the receiver operating characteristic curve (AUC) and calibration plots (slope and intercept). To identify potential fairness issues for specific demographic groups, calibration slope and intercept were also compared across gender and race/ethnicity.

3. Results

A total of 437 (3%) of 16,159 patients received a depression diagnosis within one month after the start of cancer treatment. The best performing ML model (LASSO logistic regression based on structured data) had an AUC of 0.74 (95% CI 0.71-0.78), whereas the model based solely on clinician notes performed poorly (0.50 AUC, 95% CI 0.49-0.52). The model underestimated risks for female and Non-Hispanic Black patients and overestimated for male and Non-Hispanic Asian patients.

4. Discussion

The best performing model (LASSO logistic regression on structured data) had reasonable AUC and calibration. We found discrepant model calibration across race/ethnicity and sex. The miscalibration could result in a disproportionate amount of missed patients needing additional mental health resources in specific groups. A next step could be to apply bias mitigation techniques for in- or post-processing during model development, like threshold selection and recalibration within groups.

5. Conclusion

We developed a robust model to predict depression risk among patients with cancer and demonstrated the importance of structured data to predict depression risk. Future studies might improve the prediction of depression risk in patients with cancer by refining the outcome label and supplementing predictors with patient-reported outcome measures and social determinants of health.

References

[1] Pitman A, Suleman S, Hyde N, Hodgkiss A. Depression and anxiety in patients with cancer. BMJ. 2018;361:k1415. doi: 10.1136/bmj.k1415.

[2] Dreismann L, Goretzki A, Ginger V, Zimmermann T. What if… I Asked Cancer Patients About Psychological Distress? Barriers in Psycho-Oncological Screening From the Perspective of Nurses-A Qualitative Analysis. Front Psychiatry. 2021;12:786691. Epub 2022/02/15. doi: 10.3389/fpsyt.2021.786691. PubMed PMID: 35153856; PubMed Central PMCID: PMCPMC8825354.

[3] Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:191001108. 2019.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230275

Few-Shot and Prompt Training for Text Classification in German Doctor's Letters

Phillip RICHTER-PECHANSKI[a,b,c,d,e], Philipp WIESENBACH[a,b,d], Dominic M. SCHWAB[b], Christina KIRIAKOU[b], Mingyang HE[a,b,e], Nicolas A. GEIS[b,d], Anette FRANK[e*] and Christoph DIETERICH[a,b,c,d*1]
[a] *Klaus Tschira Institute for Computational Cardiology, Heidelberg, Germany*
[b] *Department of Internal Medicine III, University Hospital Heidelberg, Germany*
[c] *German Center for Cardiovascular Research (DZHK) - Partner site Heidelberg/Mannheim, Germany*
[d] *Informatics for Life, Heidelberg, Germany*
[e] *Department of Computational Linguistics, Heidelberg University, Germany*

Abstract. To classify sentences in cardiovascular German doctor's letters into eleven section categories, we used pattern-exploiting training, a prompt-based method for text classification in few-shot learning scenarios (20, 50 and 100 instances per class) using language models with various pre-training approaches evaluated on CARDIO:DE, a freely available German clinical routine corpus. Prompting improves results by 5-28% accuracy compared to traditional methods, reducing manual annotation efforts and computational costs in a clinical setting.

Keywords. deep learning, prompting, language models, cardiology

1. Introduction and Methods

By using methods of natural language processing (NLP) and machine learning (ML) we aim to extract clinical information from unstructured doctor's letters. While most supervised ML approaches rely on large amounts of manually annotated training data, recent developments in NLP showed promising results in text classification tasks using pre-trained language models (PLM) and prompts [1]. Prompting exploits the ability of PLMs to infer knowledge from context, in combination with supervised methods they achieve state-of-the-art results on various text classification tasks. Doctor's letters are separated into sections, e.g. anamnesis, diagnosis and risk factors, which contain semantically related sentences. Here we present our initial results using pattern-exploiting training (PET) with various domain and task-adapted PLMs [2,3] on the task of section classification in German doctor's letters from the cardiology domain.
Our data is based on the CARDIO:DE corpus, a freely available and distributable large German clinical corpus from the cardiovascular domain encompassing 500 clinical routine German doctor's letters from Heidelberg University Hospital

[1] *Corresponding authors: Phillip RICHTER-PECHANSKI, Christoph DIETERICH E-mails: phillip.richter-pechanski@med.uni-heidelberg.de , christoph.dieterich@med.uni-heidelberg.de, frank@cl.uni-heidelberg.de. CD and AF share the last authorship.

(https://doi.org/10.11588/data/AFYQDY). For evaluation we used eleven CDA-compliant section categories included in CARDIO:DE such as medication and anamnesis.

We evaluated four medium-sized PLMs based on bidirectional transformer-based BERT encoder models [4]: (1) based on a publicly available German BERT model (*gbert-base*), (2) further task-adapted on our CARDIO:DE data set (*gbert-base-cardiode*), (3) further domain-adapted on an internal medical data set (200,000 German cardiology doctor's letters; *gbert-fine*), and (4) the two latter approaches combined (*gbert-fine-cardiode*). To evaluate prompting in a few-shot learning scenario we used PET, a state-of-the-art semi-supervised few-shot learning method using prompts, to classify sequences of text, achieving promising results in various domains [5]. We evaluated PET on three different training set sizes |T| = 20, 50, 100 and compared PET results with a baseline based on BERT with a sequence classification head.

2. Results

Table 1. Mean accuracy scores for |T| = 20, 50, 100 for four PLMs using a traditional sequence classification (SC) baseline model and PET. Trained on two different random seeds and training sets randomly extracted from CARDIO:DE400. We evaluated the models on 13,563 separate section annotations of CARDIO:DE100.

Model	SC |T|=20	PET |T|=20	SC |T|=50	PET |T|=50	SC |T|=100	PET |T|=100
gbert-base	28,2	54,7	45,3	67,3	62,9	72,3
gbert-base-cardiode	32,6	57,7	58,7	70,4	70	75,1
gbert-fine	37,1	57,6	60,2	70	71,7	76,4
gbert-fine-cardiode	28,4	**64,2**	48,1	**76**	67,5	**79,4**

3. Discussion and Conclusion

PET outperforms all SC baselines by large margins using any type of PLMs. The domain- and task-adapted PLM *gbert-fine-cardiode* outperforms the baseline and all other PET results. Already *gbert-cardiode* outperforms *gbert-base*, for both the baseline and PET.

PET can significantly improve classification results in a clinical setup on low-resource languages like German and can both accelerate and improve the development of accurate section classification models to e.g. support automatic medication extraction.

References

[1] P. Liu, W. Yuan, Z. Jiang, H. Hayashi, G. Neubig, J. Fu, W. Yuan, Z. Jiang, H. Hayashi, G. Neubig, and J. Fu, Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, ACM Comput Surv. 55 (2023) 1–35. doi:10.1145/3560815.

[2] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N.A. Smith, Don't Stop Pretraining: Adapt Language Models to Domains and Tasks, in: ACL, 2020: pp. 8342–8360. doi:10.18653/v1/2020.acl-main.740.

[3] T. Schick, and H. Schütze, Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference, in: EACL 2021 , 2021: pp. 255–269. doi:10.18653/v1/2021.eacl-main.20.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, (2018). http://arxiv.org/abs/1810.04805 (Mar. 7, 2019).

[5] T. Schick, and H. Schütze, True Few-Shot Learning with Prompts -- A Real-World Perspective, (2021). https://arxiv.org/abs/2111.13440v1 (Dec. 14, 2021).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230276

In-Hospital Cancer Mortality Prediction by Multimodal Learning of Non-English Clinical Texts

Shintaro OYAMA[a1], Taiki FURUKAWA[b], Shotaro MISAWA[c], Ryuji KANO[c], Hirokazu YARIMIZU[c], Tomoki TANIGUCHI[c], Kohei ONODA[c], Kikue SATO[b] and Yoshimune SHIRATORI[b]

[a] *Innovative Research Center for Preventive Medical Engineering, Nagoya University, Nagoya, Japan*

[b] *Nagoya University Hospital Medical IT Center, Nagoya University, Nagoya, Japan*

[c] *FUJIFILM Corporation, Tokyo, Japan*

Abstract. Predicting important outcomes in patients with complex medical conditions using multimodal electronic medical records remains challenge. We trained a machine learning model to predict the inpatient prognosis of cancer patients using EMR data with Japanese clinical text records, which has been considered difficult due to its high context. We confirmed high accuracy of the mortality prediction model using clinical text in addition to other clinical data, suggesting applicability of this method to cancer.

Keywords. Non-English EMR clinical text processing, Inpatient Cancer Mortality Prediction

1. Introduction and Methods

Inpatient outcome and admission prediction has been limited to the use of generalized and simple clinical decision rules (CDRs), which are not sufficiently accurate for individualized prediction, especially for complex pathologies such as those often seen in cancer patients. Using machine learning (ML) techniques that utilizes the multiple variables obtained from electronic medical record (EMR) could solve these problems and may lead to advanced clinical decision support systems. Previous studies have reported mortality prediction of pneumonia patients [1] and general inpatients prediction [2]. However, these studies were limited to specific clinical conditions and did not use the data from medical records, which presented a challenge to accuracy. In particular, medical records entered by physicians and nurses have been considered to have a high importance on supplement subtle changes in a patient's condition and contribute to accurate prediction [3], but they are mostly unstructured data, which makes them difficult to use. In addition, research on natural language processing in medical records has been conducted in English and other low-content languages, and not enough has been done in high-context languages such as Japanese [4]. In this study, we trained ML model to predict in-hospital mortality of cancer patients using multimodal clinical information

[1] Corresponding Author: Shintaro Oyama, E-mail: oyama.shintaro.m5@f.mail.nagoya-u.ac.jp.

including Japanese clinical texts, and tested the accuracy of the prediction and the difference in accuracy by cancer tumor. EMR data were extracted for patients admitted to a single tertiary hospital (Nagoya University Hospital) from January 2018 to November 2020. The data characteristics include total admissions: 55,687; mean age: 62.15yo; mortality rate: 1.15%. Inpatients admitted after January 2020 were treated as test data and the rest as training data. Since the number of deaths (205 cases) and survivors (20158 cases) differed greatly in the training data, we down-sampled survivors to 10 times the number of deaths, as an input feature, we extracted medical statements, tests, procedures, fluid and food intake, and basic patient information as input variables from EMR data up to the second day of admission, and used the Gradient Boosting Decision Trees (GBDT)-based model that we have constructed in a previous study [1]. Test data were selected for cancer types (lung cancer, hepatocellular carcinoma, and colorectal cancer) with more than fifty patients including at least three death cases during the period, and predictive performance was evaluated using the Area Under the ROC Curve (AUROC score) for the three cancer types along with all diseases and cancer types.

2. Results, Discussion and Conclusions

Table 1. represents the statistics of our test datasets, and the performance of cancer mortality prediction of our model was competitive to that of all diseases. In addition, mortality of lung cancer was the most difficult to predict.

Table 1. Statistics and Performance on our test datasets. ADM:admission

Disease Type	#ADM	Mean days from ADM. to death	Mortality rate	AUROC
All-Diseases	10,159	12	1.17%	96.50
All-types of cancer	2,202	12	1.95%	97.07
Lung cancer	374	16	3.21%	90.24
Hepatocellular carcinoma	184	6	1.63%	98.90
Colon cancer	54	17	3.70%	92.31

Predicting inpatient outcomes for complex diseases such as cancer has been considered difficult [2], thus improving accuracy, especially by using unstructured medical text in highly contextual languages such as Japanese, has been sought both for proper medical resource allocation and as a benchmark for hospital quality improvement. We constructed a multimodal machine learning model to predict inpatient mortality from Japanese EMRs and were able to predict inpatient mortality for cancer as well as other diseases. However, this is a preliminary study conducted on a dataset from a single medical institution and needs to be validated with medical information from multiple medical institutions.

References

[1] Furukawa T, Oyama S, Sato K, Misawa S, Kano R, Yarimizu H, Shiratori Y. Development of AI models to predict mortality and long-term hospitalization in pneumonia.
[2] Soffer S, et al. Predicting in-hospital mortality at admission to the medical ward: a big-data machine learning model. The American journal of medicine. 2021 Feb 1;134(2):227-34.
[3] Sheikhalishahi S, et al. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR medical informatics. 2019 May 7;7(2):e12239.
[4] Névéol A, et al. Clinical natural language processing in languages other than English: opportunities and challenges. Journal of biomedical semantics. 2018 Dec;9(1):1-3.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230277

Translating Medical Dialogues into Pictographs: An Approach Using UMLS

Johanna GERLACH[a,1], Pierrette BOUILLON[a], Magali NORRÉ[a,b] and Hervé SPECHBACH[c]
[a] *FTI/TIM, University of Geneva, Geneva, Switzerland*
[b] *CENTAL, ILC, Catholic University of Louvain, Louvain-la-Neuve, Belgium*
[c] *HUG, Geneva University Hospitals, Geneva, Switzerland*

Abstract. This paper describes a first attempt to map UMLS concepts to pictographs as a resource for translation systems for the medical domain. An evaluation of pictographs from two freely available sets shows that for many concepts no pictograph could be found and that word-based lookup is inadequate for this task.

Keywords. Pictographs, medical communication, umls

1. Introduction

Including pictographs in medical communication has been shown to improve patient comprehension [1]. Different systems exist to map text into pictographs, e.g., PictoBert, a general domain word-sense language representation model that predicts pictographs using WordNet and the ARASAAC pictograph database. In the context of the PROPICTO project, we are developing PictoDr, a translation system for French medical dialogues. Instead of a word-based approach, where homographs and medical multiword expressions would lead to errors, we chose a concept-based approach. PictoDr proceeds in two steps: 1. translating spoken utterances into a UMLS gloss (sequence of Unified Medical Language System (UMLS) concepts [2]) using a neural classification approach [3], and 2. mapping these concepts to pictographs. In this paper, we describe the creation of resources for this second step, using pictographs from ARASAAC and SantéBD[2].

2. Methods

The UMLS concepts were extracted from the PictoDr training data which consist of 10k diagnostic questions and medical instructions, each linked to many French variations and a UMLS gloss. The resulting 1656 unique concepts were mapped to pictographs, based on associated words for ARASAAC and on filenames for SantéBD, resulting in 1182 concepts with between 1 and 54 pictograph candidates each, for a total of 3655 pictographs. Three doctors evaluated the usefulness of these pictographs in the context of the PictoDr sentences, rating them on a five-point scale from “useful” to “useless”.

[1] Corresponding Author: Johanna Gerlach, E-mail: johanna.gerlach@unige.ch
[2] https://arasaac.org/ and https://santebd.org/

3. Results

Of the 3655 included pictographs, 983 were found to be useful representations of the UMLS concept by at least 2 of the 3 evaluators. Table 1 shows the concepts grouped by UMLS semantic type with the number of concepts for which one or more of the pictographs was rated useful. Physical objects (e.g. medical devices, anatomical structures and substances) were the most successful with 79% obtaining an illustration, followed by activities (e.g. diagnostic procedures) with 74%. Conceptual entities (e.g. findings, signs and symptoms, qualitative concepts), which represent more than half of the studied concepts, as well as phenomena or processes (e.g. diseases) were less successful with only 60% finding a useful illustration. Overall 2/3 of the included concepts obtained one or more useful pictographs (ARASAAC and/or SantéBD).

Table 1. Concepts with pictographs found useful by at least 2 doctors, grouped by UMLS semantic type.

Semantic type	Concepts	# with at least 1 useful pictograph	%
(A1) Physical Object	208	164	79%
(A2) Conceptual Entity	647	391	60%
(B1) Activity	158	117	74%
(B2) Phenomenon or Process	144	88	61%
other (not UMLS)	25	15	60%
Total	1182	775	66%

4. Discussion and Conclusions

Close to three quarters of the PictoDr concepts obtained at least one pictograph candidate. However, in the case of SantéBD, identifying candidates based on filenames which were not designed for this task led to many potentially useful images being skipped.

Among the candidates, pictographs were rejected for different reasons. Some were marked as too specific, i.e., including more information than the concept, for example a wound on a specific body part rather than a wound alone for the concept “Injury wounds”, thus making them unsuitable for generic use in a sequential representation as proposed in PictoDr. Initial comprehensibility evaluations have shown that such pictographs can be confusing depending on the context. Others were rejected because they were mapped incorrectly due to ambiguous pictograph filenames, e.g. “strong” which in the general domain could be illustrated with a physically strong person, but is not appropriate in the context of “strong pain”.

References

[1] Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. Patient Education and Counseling. 2006; 61(2):173–190.

[2] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D267-70. doi: 10.1093/nar/gkh061.

[3] Mutal J, Bouillon, P, Norré M, Gerlach J, Ormaechea Grijalba, L. A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation System - The BabelDr Use Case, Proceedings of AMTA, 2022.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230278

SapBERT-Based Medical Concept Normalization Using SNOMED CT

Akhila ABDULNAZAR[a,b,1], Markus KREUZTHALER[a], Roland ROLLER[c] and Stefan SCHULZ[a]
[a] *Institute for Medical Informatics Statistics and Documentation, Medical University of Graz, Austria*
[b] *Center for Biomarker Research in Medicine, Graz, Austria*
[c] *German Research Center for Artificial Intelligence, Germany*

Abstract. Word vector representations, known as embeddings, are commonly used for natural language processing. Particularly, contextualized representations have been very successful recently. In this work, we analyze the impact of contextualized and non-contextualized embeddings for medical concept normalization, mapping clinical terms via a k-NN approach to SNOMED CT. The non-contextualized concept mapping resulted in a much better performance (F_1-score = 0.853) than the contextualized representation (F_1-score = 0.322).

Keywords. Medical Concept Mapping, SNOMED CT

1. Introduction

In clinical language processing, BERT-based models have exceled by considering discourse context. So should a vector representation of the word *"cold"* be capable to disambiguate whether the referent is a disease or a temperature. Mapping spans of words in clinical narratives to codes from terminologies such as SNOMED CT can be achieved through similarity matching in an n-dimensional embedding space. Taking context, i.e., the surrounding words into account, it should additionally be expected that the assignment of the correct code to an ambiguous expression is supported.

2. Method

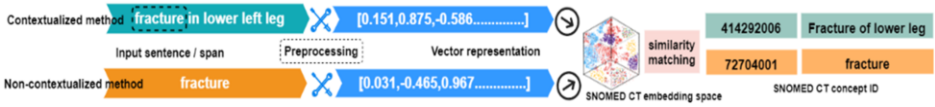


Figure 1. Contextualized and non-contextualized normalization using similarity matching from embeddings.

We used SapBERT [1], a transformer-based language model fine-tuned on UMLS with data from the 2019 n2c2 [2] normalization task. This shared task was addressed by different approaches for concept mapping, from dictionary matching to deep learning.

[1] Corresponding Author: Akhila Abdulnazar, E-mail: akhila.kuppassery-abdulnazar@stud.medunigraz.at

Our goal was to reuse this framework in order to compare contextualized with non-contextualized mapping of text spans to SNOMED CT. First, the English synonyms from the international SNOMED CT version were mapped to vector representations using SapBERT. Second, concept mentions in sentences of the annotated n2c2 dataset were vectorized in the same way. De-capitalization, stop word and special character removal had been performed in a preprocessing step.

We compared (i) text mentions in their local context (a line of text) with (ii) the text mentions alone. SNOMED CT term candidates were retrieved by a nearest neighbor's search based on cosine similarity, as detailed in Figure 1, and finally mapped to the SNOMED CT code they belonged to.

3. Results and Conclusion

The non-contextualized representation showed a much higher performance, (Table 1), with an F_1-score more than twice as high. This illustrates the lack of context information within the embedding space and exemplifies the necessity of contextualization of SNOMED CT concepts and synonyms, usually not available in official terminology releases. This makes a generic contextualized medical concept normalization approach for about ~350k concepts not feasible at the moment.

Table 1. Evaluation results of contextualized and non-contextualized concept mapping on the n2c2 dataset.

Method	Precision	Recall	F_1-score
Contextualized representation	0.492	0.279	0.322
Non-contextualized representation	0.870	0.847	0.853

An error analysis on the non-contextualized approach revealed certain types of errors. The non-contiguous nature of the spans results in contextual errors. The non-uniformity in providing one 'correct' normalization to a span returns analogy and granularity errors. The least occurring kind of errors resulted from spelling errors and acronyms.

In conclusion, the k-NN approach via non-contextual concept representations leveraging SapBERT on an existing concept normalization dataset scored better in comparison to the degrading performance of the contextual representation of the normalization candidate. Future investigations will focus on a generic unsupervised medical concept normalization approach of SNOMED CT aiming for a combination of contextualized with non-contextualized representation schemes. This will require enhancing the contexts of SNOMED CT concepts with real-world clinical information.

References

[1] Liu F, Shareghi E, Meng Z, Basaldella M, Collier N. Self-alignment pretraining for biomedical entity representations. arXiv preprint arXiv:2010.11784. 2020 Oct 22. doi: 10.18653/v1/2021.naacl-main.334

[2] Luo YF, Henry S, Wang Y, Shen F, Uzuner O, Rumshisky A. The 2019 n2c2/UMass Lowell shared task on clinical concept normalization. Journal of the American Medical Informatics Association. 2020 Oct;27(10):1529-e1. doi: 10.1093/jamia/ocaa106.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230279

Data-Driven Identification of Clinical Real-World Expressions Linked to ICD

Amila KUGIC[a,1], Bastian PFEIFER[a], Stefan SCHULZ[a] and Markus KREUZTHALER[a]
[a] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract. A semi-structured clinical problem list containing ~1.9 million de-identified entries linked to ICD-10 codes was used to identify closely related real-world expressions. A log-likelihood based co-occurrence analysis generated seed-terms, which were integrated as part of a k-NN search, by leveraging SapBERT for the generation of an embedding representation.

Keywords. Natural Language Processing, Big Data, Electronic Health Records

1. Introduction

Electronic health records are largely constituted by non-standard narratives, which contain various domain-typical expressions that are not well represented by current medical terminology systems. Annotated resources of clinical real-world data show promising results as they can be leveraged to support data-driven term candidate generation to enrich terminology systems with synonyms. Thus, they are better adapted to clinical jargon in a given natural language, which is particularly useful for clinical entity normalization using natural language processing.

In biomedical terminology expansion, co-occurrence analyses and embeddings-based representation schemes have separately been used to harvest term candidates. Examples of these include the integration of hierarchical information into the ICD-10 coding standard [1] and fine-tuning word-vector pairs to generate suitable synonyms [2]. In combining both approaches, this investigation aimed to repurpose ICD-10-coded problem list entries to identify term candidates in a semi-supervised way. Previous work [3] focused on applying co-occurrence analysis using a log-likelihood based ranking.

2. Data

The German dataset from an Austrian hospital provider has ~1.9 million unique de-identified problem list items, entered by physicians, in conjunction with assigned ICD-10 codes for administrative purposes. The items exhibit typical features of clinical language, such as abbreviations, acronyms, misspellings and non-standardized numeric expressions, e.g.: “Diab. mellitust Typ 2, HbA1c: 43 mmol/mol”.

[1] Corresponding Author: Amila Kugic, E-mail: amila.kugic@medunigraz.at

3. Methods

Preprocessing. We normalized the text entries by removing any characters except [a-zA-Z0-9üäöÄÖÜß-], followed by white space tokenization. For log-likelihood ratios, Apache Spark and the MapReduce programming were applied, as well as indexing co-occurrence analysis results with Apache Lucene.
Seed-term generation. An n-gram, paired with an ICD code, with a significant $p<0.01$ log-likelihood value was considered a valid seed-term candidate.
Embedding space generation. Per unique n-gram decomposition, we built a 768-dimensional embedding space indexing the resulting vectors via Faiss [4]. SapBERT [5] was applied without downstreaming due to the highly imbalanced dataset.
Search strategy. The seed term candidates were utilized to perform a k-NN search using the L2 distance without the square root due to performance issues. The top 10 candidates were filtered according to two criteria: a weighted nearest neighbor distance and a syntactical filter.
Evaluation strategy. "I25.3 – Aneurysm of heart" was selected for evaluating the described methodology.

4. Results and Outlook

With the co-occurrence analysis alone, two exact synonyms and eight hyponyms could be identified. Using these identified term candidates for a k-NN search in the embedding space, in addition two synonyms (e.g., "Herzwandaneurysma"), seven hyponyms (e.g., "Herzventrikelaneurysma") and one incorrect candidate could be extracted. Moreover, by utilizing the standardized description from the German ICD-10 release as seed-term for the embedding space-based k-NN approach, further two synonyms (e.g., "Herz-Aneurysma") and eight hyponyms (e.g., "Herzspitzen-Aneurysma") were harvested. In both approaches, acronyms as part of word compounds and syntactical variants of the original seed term could be identified.

This scenario showcases a first attempt to combine a co-occurrence analysis with an embeddings-based k-NN approach in order to find new term candidates not present in clinical terminology systems. Future work will investigate the influence of a downstreamed language model to this problem domain, and evaluations will branch out to include a variety of different ICD-10 codes.

References

[1] Finch A, Crowell A, Bhatia M, Parameshwarappa P, Chang YC, Martinez J, Horberg M. Exploiting hierarchy in medical concept embedding. JAMIA Open. 2021 Jan 1;4(1):ooab022.
[2] Gu G, Zhang X, Zhu X, Jian Z, Chen K, Wen D, et al. Development of a consumer health vocabulary by mining health forum texts based on word embedding: semiautomatic approach. JMIR medical informatics. 2019;7(2):e12704.
[3] Kreuzthaler M, Pfeifer B, Schulz S. Terminology Expansion via Co-occurrence Analysis of Large Clinical Real-World Datasets. In: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). IEEE; 2022. p. 01–2.
[4] Johnson J, Douze M, Jégou H. Billion-scale similarity search with gpus. IEEE Transactions on Big Data. 2019;7(3):535–47.
[5] Liu F, Shareghi E, Meng Z, Basaldella M, Collier N. Self-alignment pretraining for biomedical entity representations. arXiv preprint arXiv:201011784. 2020

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230280

Digital Health Data Capture with a Controlled Natural Language

Kristian Juha Ismo KANKAINEN[a,1], Inari LISTENMAA[b], Gunnar PIHO[a] and Peeter ROSS[a]
[a] *Department of Health Technologies, Tallinn University of Technology, Estonia*
[b] *Digital Grammars AB, Sweden*
ORCiD ID: Kristian Juha Ismo Kankainen https://orcid.org/0000-0002-0551-927X,
Inari Listenmaa https://orcid.org/0000-0001-5555-0251,
Gunnar Piho https://orcid.org/0000-0003-4488-3389,
Peeter Ross https://orcid.org/0000-0003-1072-7249

Abstract. Written text has been the preferred medium for storing health data ever since Hippocrates, and the medical narrative is what enables a humanized clinical relationship. Can't we admit natural language as a user-accepted technology that has stood against the test of time? We have previously presented a controlled natural language as a human-computer interface for semantic data capture already at the point of care. Our computable language was driven by a linguistic interpretation of the conceptual model of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). This paper presents an extension that allows the capture of measurement results with numerical values and units. We discuss the relation our method can have with emerging clinical information modelling.

Keywords. Data capture, controlled natural language, semantic health data, SNOMED CT, information model

1. Introduction and Method

Modern clinical terminologies are undoubtedly expressive, essentially due to compositionality by post-coordination. For example, the SNOMED CT includes 432 person concepts and 52,181 causative agent concepts (as of release 2022-11-30). Consequently, the two concept groups can be composed to form more than 22 ½ million post-coordinated expressions. Nonetheless, a recent scoping literature review concluded that this expressiveness is rarely used and that "there is no easy solution for mapping free text to this terminology and to perform automatic post-coordination" [1].

We consider this lack of mapping to be due to the lack of a good user interface and have earlier proposed (detailed in [2]) a novel mapping in the form of a computable controlled natural language (CNL) that corresponds to a linguistic interpretation of SNOMED CT's *Situation With Explicit Context*. We insist a language-based user interface is appropriate for data capture as the written text has been the medium for health data ever since Hippocrates, and language is what enables a humanized clinical relationship [3].

[1] Corresponding Author: Kristian Juha Ismo Kankainen, E-mail: kristian.kankainen@taltech.ee.

We use Grammatical Framework (GF) [4] to program our CNL. Our GF grammar consists of one abstract application grammar and concrete syntaxes for Estonian, English, and SNOMED CT compositional grammar. The rule-based approach of GF supports explainability [5], which we see as crucial in healthcare.

2. Results and Discussion

A new concept *ObservedCondition* was added to our abstract grammar. This new type takes a numerical quantity, an evaluation procedure, and an observable entity as input parameters. Both latter are ordinary SNOMED CT objects. Although SNOMED CT has introduced support for numeric values for medications [6], it is not able to express numerical measurement results. As we cannot map numerals to SNOMED CT, we have instead started to explore mapping to clinical information models. Furthermore, our extended abstract model now complies with ISO 13940 ContSys as "a number of healthcare investigations reveal a number of observed conditions" [7].

Mapping to clinical information models makes sense to us in two ways. Grounding the necessary modeling work on the language used for data capture, can enable the consensual development of reusable domain information models [8]. As an example, our presented extension employs information about the unit encoded in the observable entity object. Although this encoding follows the Snomed CT concept model, it is little used, as only 23 terms specify a unit in the 2022-11-30 international release. Arguably, contextual information, like unit, could be managed in a national extension or repository.

Accommodating the data capturer's view can also improve user experience [9]: GF enables actionable text (data) entry. As GF translation functions in both ways, it can render already captured data as text, which can further allow multilingual authoring of the same health record by several actors across healthcare borders.

References

[1] Gaudet-Blavignac C, Foufi V, Bjelogrlic M, Lovis C. Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for Processing Free Text in Health Care: Systematic Scoping Review. Journal of Medical Internet Research. 2021 Jan;23.

[2] Kankainen K, Klementi T, Piho G, Ross P. Using SNOMED CT as a Semantic Model for Controlled Natural Language Guided Capture of Clinical Data. CEUR Workshop Proceedings. 2022 Aug;Vol-3264.

[3] Miranda-Bastidas C. Medical records: the medical writing of the patient's story to the medical narrative. Colombia Medica. 2020 Jun;51.

[4] Ranta A. Grammatical framework: programming with multilingual grammars. Studies in computational linguistics. CSLI Publications; 2011.

[5] Ranta A. Explainable machine translation with interlingual trees as certificates. In: Proceedings of the Conference on Logic and Machine Learning in Natural Language (LaML 2017); 2017. p. 63-78.

[6] SNOMED International. SNOMED International delivers technical preview for enhancement of SNOMED CT with concrete domains in advance. 2021. Available from: https://www.snomed.org/news-and-events/articles/snomed-delivers-technical-preview-concrete-domains.

[7] ISO 13940:2015(en), Health informatics — System of concepts to support continuity of care.

[8] Oemig F, Blobel B. The Value of Domain Information Models for Achieving Interoperability. pHealth 2020. 2020:75-82. IOS Press.

[9] Jansson M, Liisanantti J, Ala-Kokko T, Reponen J. The negative impact of interface design, customizability, inefficiency, malfunctions, and information retrieval on user experience: A national usability survey of ICU clinical information systems in Finland. International Journal of Medical Informatics. 2022 Mar;159.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230281

Information Extraction from Medical Texts with BERT Using Human-in-the-Loop Labeling

Hendrik ŠUVALOV[a,1] and Sven LAUR[a] and Raivo KOLDE[a]
[a] *University of Tartu, Estonia*
ORCiD ID: Hendrik Šuvalov https://orcid.org/0000-0001-9625-3552,
Sven Laur https://orcid.org/0000-0002-9891-3347,
Raivo Kolde https://orcid.org/0000-0001-9625-3552

Abstract. Neural network language models, such as BERT, can be used for information extraction from medical texts with unstructured free text. These models can be pre-trained on a large corpus to learn the language and characteristics of the relevant domain and then fine-tuned with labeled data for a specific task. We propose a pipeline using human-in-the-loop labeling to create annotated data for Estonian healthcare information extraction. This method is particularly useful for low-resource languages and is more accessible to those in the medical field than rule-based methods like regular expressions.

Keywords. BERT, information extraction, natural language processing, medical texts, named entity recognition

1. Introduction

Low-resource languages, such as Estonian, do not have much available annotated data which makes training neural network language models difficult [1]. We use human-in-the-loop labeling to annotate data for finding specific entities like drug names, disease names or classifications etc. [2]. First, we use a naive regex with as high recall as possible to get samples with more frequent positive examples, then we start annotating the data, train the model and use its predictions to annotate the rest of the data faster. We used this method to fine-tune a model pre-trained on unlabeled Estonian medical data to extract cancer *TNM* stages[2] from free text and got remarkable results.

2. Methods

We used a naive regex to enrich the initial texts with positive examples of *TNM* stages. We then started annotating the data and after every 50 annotations, we used a BERT model [3] pre-trained on Estonian electronic health records and fine-tuned it with all the labeled data. We then used its predictions to find the positive examples to annotate faster.

[1] Corresponding Author: Hendrik Šuvalov, E-mail: hendrik.suvalov@ut.ee.
[2] *TNM* stages - https://www.cancer.gov/about-cancer/diagnosis-staging/staging

With each new iteration of training, we saw how accurately it predicted the stages and stopped when we were convinced it had learned the task.

3. Results

We evaluated the performance of the model, taking our advanced regular expression that we use in our workflows as the ground truth. Our test set consisted of 10'000 examples with 1:10 positive example ratio. After the model was trained on 150 examples, among which 14 were positive examples, it started predicting *TNM* stages with precision of 0.914 and recall of 0.867 and after 500 examples, it achieved precision of 0.820 and recall of 0.951. We manually analyzed the false positives annotated by the model of which there were 200 and found that 38 of them were actually correct, meaning the BERT based tagger caught the cases, but our regular expressions did not. The code is available at https://github.com/HealthInformaticsUT/labelstudio-ner-tagger.

4. Discussion

Extracting *TNM* stages is a relatively simple task that does not require many annotations for the model to learn and is not very context-dependent, which is one of BERT's strengths in comparison to rule-based approaches [3]. For more ambiguous cases, such as extracting mentions of family history, it could be much more efficient. Also, selecting which instances to train with or annotate can be approached in many ways, for example clustering similar texts or correcting the annotations for cases where the model is not very confident [4].

5. Conclusions

Pre-trainable neural network language models are useful for information extraction tasks but require annotated data, which low-resource languages often lack [1]. Human-in-the-loop labeling is an effective method for generating these annotations and the resulting models trained on this data can produce impressive results, often extracting cases that commonly used rule-based methods miss [2].

References

[1] Liu Z, Jiang F, Hu Y, Shi C, Fung P. NER-BERT: A pre-trained model for low-resource entity tagging. arXiv [csCL] [Internet]. 2021 [cited 2023 Jan 9], doi: https://doi.org/10.48550/arXiv.2112.00405

[2] Zhao Y, Liu J. Human-in-the-loop based named entity recognition. In: 2021 International Conference on Big Data Engineering and Education (BDEE). IEEE; 2021. p. 170–6, doi: https://doi.org/10.1109/BDEE52938.2021.00037

[3] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. arXiv [csCL] [Internet]. 2018 [cited 2023 Jan 10], doi: https://doi.org/10.48550/arXiv.1810.04805

[4] Zhang Y, Nie A, Zehnder A, Page RL, Zou J. VetTag: improving automated veterinary diagnosis coding via large-scale language modeling. NPJ Digit Med [Internet]. 2019;2(1):35

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230282

Large Language Model as Unsupervised Health Information Retriever

Keyuan JIANG[a,1], Mohammed M. MUJTABA[a] and Gordon R. BERNARD[b]
[a] *Purdue University Northwest, Hammond, Indiana, USA*
[b] *Vanderbilt University, Nashville, Tennessee, USA*
ORCiD ID: Keyuan Jiang https://orcid.org/0000-0002-1565-3202

Abstract. Retrieving health information is a task of search for health-related information from a variety of sources. Gathering self-reported health information may help enrich the knowledge body of the disease and its symptoms. We investigated retrieving symptom mentions in COVID-19-related Twitter posts with a pretrained large language model (GPT-3) without providing any examples (zero-shot learning). We introduced a new performance measure of total match (TM) to include exact, partial and semantic matches. Our results show that the zero-shot approach is a powerful method without the need to annotate any data, and it can assist in generating instances for few-shot learning which may achieve better performance.

Keywords. Large language model, unsupervised learning, zero-shot learning, health information retrieval, COVID-19 symptoms, Twitter

1. Introduction

Recent advancements in pre-trained large language models (LLMs) have revolutionized the ways of natural language processing. They achieved state of the art (SOTA) performances in many NLP tasks including conversation, text generation, reasoning, translation, and question answering. To retrieve the symptom mentions from a corpus of COVID-19-related tweets, we treat the retrieval as a question answering task. This treatment facilitated selecting the prompt to the language model and measuring the performance of the unsupervised (zero-shot) learning.

2. Method

The 3rd generation pre-trained transformer (GPT-3) language model [1], developed by Open AI, is an autoregressive language model developed based upon the previous generation (GPT-2) [2]. We used the GPT-3 model (text-davinci-002) without modifications of its architecture and hyperparameters for zero-shot and few-shot learning (5-shot and 10-shot)[2].

[1] Corresponding Author: Keyuan Jiang, Purdue University Northwest, 2200 169th Street, Hammond, Indiana, U.S.A., E-mail: kjiang@pnw.edu.
[2] The work of the few-shot learning methods is under review at another conference.

Choosing an appropriate prompt to the language model is an important step. We adopted the prompt of "List the symptoms in this tweet as a list" after having experimented several different prompts. We proposed the total match (TM) for the practice use cases, rather than the widely used, strict exact match (EM) [3]. TM includes exact, partial and semantical matches. A corpus of 655 COVID-19-related tweets, annotated by two authors, was used to test the zero-shot learning.

3. Results and Discussions

Table 1 are the performance results of the zero-shot method along with the results of both 5-shot and 10-shot methods using GPT-3 on a corpus of 655 annotated tweets.

Table 1. Performance results of zero-shot method in comparison with 5-shot and 10-shot learning methods on 655 tweets.

	Exact Match	Total Match	Weighted TM	Precision	Recall	F1
0-shot	0.109	0.212	0.790	0.779	0.856	0.816
5-shot	0.191	0.377	0.853	0.844	0.890	0.866
10-shot	0.264	0.429	0.906	0.912	0.846	0.877

Both EM and TM are lower in our method than that of few-shot learning, suggesting that supervised learning even with a small number of examples can perform better. Precision, recall and F1 are somehow lower but comparable, showing the power of GPT-3 to recognize symptom mentions in tweets with the general domain training data. This can be useful in a larger scale health information retrieval task using few-shot learning as the zero-shot method can serve as an initial step of few-shot learning by identifying a small number of instances needed to condition the model.

It was observed that the GPT-3 language model sometimes predicts symptom mentions not found in the tweets. This may come from the facts in the data it learned. Another limitation is its inability to handle negation correctly – e.g., it predicts the phrase of "no cough" as a symptom.

4. Conclusion

The zero-shot approach of the LLM is very powerful, and can help us identify examples for few-shot learning methods when no annotated data are available, and can facilitate a larger scale task of health information retrieval.

References

[1] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877-901.

[2] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. 2019 Feb 24;1(8):9.

[3] Rajpurkar P, Zhang J, Lopyrev K, Liang P. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250. 2016 Jun 16.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230283

Announcement of the German Medical Text Corpus Project (GeMTeX)

Frank MEINEKE[a], Luise MODERSOHN[b], Markus LOEFFLER[a] and Martin BOEKER[b,1]
[a] *University of Leipzig, IMISE, Germany*
[b] *Technical University of Munich, AIIM, Germany*
ORCiD ID: Frank Meineke orcid.org/0000-0002-9256-7543,
Luise Modersohn orcid.org/0000-0002-8258-4366,
Markus Loeffler orcid.org/0000-0002-0424-9933, Co-Coordinator GeMTeX,
Martin Boeker orcid.org/0000-0003-2972-2042, Coordinator GeMTeX

Abstract. The largest publicly funded project to generate a German-language medical text corpus will start in mid-2023. GeMTeX comprises clinical texts from information systems of six university hospitals, which will be made accessible for NLP by annotation of entities and relations, which will be enhanced with additional meta-information. A strong governance provides a stable legal framework for the use of the corpus. State-of-the art NLP methods are used to build, pre-annotate and annotate the corpus and train language models. A community will be built around GeMTeX to ensure its sustainable maintenance, use, and dissemination.

Keywords. Natural Language Processing, Text Corpus, German Medical Informatics Initiative

1. Introduction

The potential of natural language processing (NLP) is quickly improving with rapid advances in machine learning esp. deep learning. The language of clinical documentation is very different in vocabulary and structure from normal written language or the language of scientific publications. Therefore, the progress of clinical NLP crucially depends on specifically trained language models, requiring authentic clinical documents. The German Medical Informatics Initiative (MII) [1] provides a unique opportunity to make clinical documents accessible on a large scale, and to enrich them with annotations. A German Medical text Corpus is expected to boost the development of NLP-resources that support German clinical text analysis [2]. GeMTeX will address two major bottlenecks that have hindered German clinical language models to date [3,4], ie. data accessibility and data annotation.

[1] Corresponding Author: Prof. Dr. Martin Boeker, Institute of Artificial Intelligence and Informatics in Medicine (AIIM), School of Medicine, Technical University of Munich, Ismaninger Str. 22, D-81675 Munich

2. Main Objectives

GeMTeX will provide more then 150K clinical texts for NLP at six German university hospitals, reflecting a wide variety of regional and clinic-specific medical language styles. Annotation is performed by teams of trained student assistants and documentalists following annotation guidelines and using state-of-the-art tools. The texts represent different medical specialties and text types in a balanced way (e.g. discharge summaries, findings reports). Meta-information is added, related structured data from the data integration centers can be accessed. Besides basic annotation, in-depth annotations is performed in four medical areas (cardiology, pathology, pharmacy, and neurology), each on a subcorpus. The annotation uses a structured annotation vocabulary based on semantic standards such as standardized terminologies (e.g., SNOMED CT, ICD-10, TNM and others), ontologies and information models (FHIR). Pre-annotations, based on existing terminologies, algorithms and models simplify and speed up the annotation process. A strong governance will adopt a scientific framework using a study protocol. Its legal regulations are based on the data protection concept and the data use regulations and contract of the MII. Access to patient data and documents is only possible where patients have agreed to the MII broad consent. Annotated documents are automatically anonymized and manually checked. Access to the document corpus for scientific projects is additionally subject to the data access regulations of the MII and must be decided by a Data Use and Access Committee at each site center. GeMTeX will provide up-to-date tools and methods for corpus generation and NLP. Depending on the research question, different integration scenarios for texts can be implemented, which allow local, distributed and central use. Distributed machine learning will be implemented within GeMTeX and in cooperation with other projects. Leading NLP companies Averbis® and ID Berlin®, participate with their respective NLP solutions in the development and use of GeMTeX.

3. Expected Results

An important advantage of GeMTeX is its open, transparent, and extensible publication, distribution, and use according to FAIR principles. Sustainability will be achieved through cooperation with the German Information Center for Life Sciences (ZB MED), which will make the generated resources (texts, methods, tools) available on a long-term basis. The corpus for German clinical texts will be accessible to scientific projects via the MII technical infrastructure and terms of use. GeMTeX will funded from mid-2023 by [1]. Due to brevity the 17 PIs of the project are not listed in the author list.

References

[1] Medical Informatics Initiative. Funding scheme of the German Federal Ministry of Education and Research (BMBF) Available from: URL: https://www.medizininformatik-initiative.de/en/start.

[2] Kittner M, Lamping M, Rieke DT, et al. Annotation and initial evaluation of a large annotated German oncological corpus. JAMIA Open. 2021;4(2):ooab025, doi:10.1093/jamiaopen/ooab025

[3] Wang J, Deng H, Liu B, et al. Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed. J Med Internet Res. 2020;22(1):e16816, doi:10.2196/16816

[4] Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical Natural Language Processing in languages other than English: opportunities and challenges. J Biomed Semantics. 2018;9(1):12. Published 2018 Mar 30. doi:10.1186/s13326-018-0179-8

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230284

Clustering Similar Diagnosis Terms

Stefan SCHULZ[1], Akhila ABDULNAZAR and Markus KREUZTHALER
IMI, Medical University of Graz, Austria

Abstract. A large clinical diagnosis list is explored with the goal to cluster syntactic variants. A string similarity heuristic is compared with a deep learning-based approach. Levenshtein distance (LD) applied to common words only (not tolerating deviations in acronyms and tokens with numerals), together with pair-wise substring expansions raised F1 to 13% above baseline (plain LD), with a maximum F1 of 0.71. In contrast, the model-based approach trained on a German medical language model did not perform better than the baseline, not exceeding an F1 value of 0.42.

Keywords. Named Entity Normalization, Electronic Health Records

1. Introduction

Clinicians prefer telegram-style expressions over controlled terms from terminologies. Instead of "Malignant neoplasm of bronchus and lung, middle lobe, bronchus or lung" (ICD-10 C34.2) they write "Adenocarcinoma, middle lobe right", or "Adeno-Ca, R middle lobe", or even "Adneocarcinoma (sic!), right middle lobe". Clinical entity normalization (CEN) should assign the same code to term variants, being tolerant regarding typos, but strict regarding lexical differences ("Vitamin A" vs. "Vitamin B", or "Hepatectomy" vs. "Hepatotomy"). Increasingly, CEN combines neural approaches with dictionaries [1]. We processed about 20.5 million short (max. 50 chars) diagnosis descriptions annotated with ICD-10. A benchmark was created of 20 random entries. The baseline, Levenshtein Similarity (LS) is based on Levenshtein distance (LD) [2]. For strings S_1 and S_2 we define: $LS(S_1, S_2) = 1 - (2 * LD(S_1, S_2) / (Length(S_1) + Length(S_2)))$.

2. Methods

We introduced SLS (Selective Levenshtein Similarity), which ignores stop words and punctuation characters (except "."). SLS requires an exact, case sensitive match for all non-standard tokens (NST), i.e., tokens with non-alpha characters or any upper-case character beyond the first position. LS is applied to the standard tokens only. Exact string match between all NST of S_1 and S_2 is required, otherwise SLS is set to zero. Thus, *SLS* ("Type 1", "Type 2") equals zero as well as *SLS* ("EEG", "ECG"). We optionally consider variants with truncated tokens such as "chron." for "chronisch" (with or without period): for each token of a string pair S_1 and S_2 (after stop word and NST removal) token

[1] Corresponding Author: S. Schulz, Medical University, Auenbruggerplatz 2/V, 8036 Graz. E-mail: stefan.schulz@medunigraz.at.

S_{1i} is substituted by S_{2i} if the former is a left-sided substring of the latter, or S_{2i} is substituted by S_{1i} in the opposite case. We also compared the original word order with alphabetic word order (AWO). This algorithmic approach is then compared to a neural method. Top matches to the vector representations of strings from an embedding space are analysed, filled with ICD terms in their vector representations obtained by downstreaming a German medical language model leveraging SapBERT [3] on random pairs (max 50 of the same ICD code) of the list, enriched by official ICD-10 terms and synonyms. Training had been done for 50 epochs on an NVIDA GeForce GTX Titan X GPU. Similarity matching was based on a k-nearest neighbour approach using Faiss [4].

Figure 1. P, R, F_1 at several points

3. Results and Discussion

Regarding F_1, the neural model did not fare better than the LS baseline whereas the algorithmic approach yielded an F_1 13% above baseline (Fig. 1). We found that the precision drop of the model-based approach was mostly due to candidates with true variants plus additional modifiers, e.g. ("Duodenalstenose" vs. "St.p. Duodenalstenose", cosine 0.92) and small but significant variations in tokens with numerals ("Spinalkanalstenose L3,L4" vs. "Spinalkanalstenose L3-L5"). Variants with abbreviations ("rez. Erbrechen" vs. "rezidivierendes Erbrechen") had a lower cosine (here 0.85). To retrieve all variant candidates, the algorithmic approach took on average 4.5 min. compared to 23 sec. of the model-based one. The explaining power of these results is limited by small sample size and data heterogeneity. The importance to apply fuzzy string matching selectively as well as the potential of the resolution of truncation-based abbreviations is emphasised. In contrast, the SapBERT model lowercases all input and does not consider abbreviations. The coarse-grainedness of ICD-10 was the reason that the model did not learn many distinctions, even that "right" and "left" are not synonyms, because they occur in nearly the same distributional contexts. Future work should emphasise combinations of the two approaches, e.g., by using the algorithmic approach to optimise the sampling for SapBERT.

References

[1] Ferré, A, Deléger L, Bossy R et al. C-Norm: a neural approach to few-shot entity normalization. BMC Bioinformatics 2020 Suppl 23, 579
[2] Levenshtein VI et al. Binary codes capable of correcting deletions, insertions, and reversals. Soviet physics doklady. 1966; S. 707–710.
[3] Liu F, Shareghi E, Meng Z, Basaldella M, Collier N. Self-Alignment Pretraining for Biomedical Entity Representations. Proc. of NACL Human Language Technologies 2021 Jun, 4228–4238.
[4] Johnson J, Douze M, Jégou H. Billion-scale similarity search with GPUs. IEEE Transactions on Big Data. 2019;7(3):535–547.

Section 8

Public Health and Epidemiology Informatics

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230286

Health Research Requires Efficient Platforms for Data Collection from Personal Devices

Erlend JOHANNESSEN[a,1], André HENRIKSEN[a], Eirik ÅRSAND[a], Alexander HORSCH[a], Jonas JOHANSSON[b] and Gunnar HARTVIGSEN[a]
[a] *Dept. of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway*
[b] *Dept. of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway*
ORCiD ID: Johannessen 0000-0003-4860-9192, Henriksen 0000-0002-0918-7444, Årsand 0000-0002-9520-1408, Horsch 0000-0001-7745-0139, Johansson 0000-0001-7912-5786, Hartvigsen 0000-0001-8771-9867

Abstract. Data from consumer-based devices for collecting personal health-related data could be useful in diagnostics and treatment. This requires a flexible and scalable software and system architecture to handle the data. This study examines the existing mSpider platform, addresses shortcomings in security and development, and suggests a full risk analysis, a more loosely coupled component-based system for long term stability, better scalability, and maintainability. The goal is to create a human digital twin platform for an operational production environment.

Keywords. Infrastructure, Scalability, Human Digital Twin

1. Introduction

Physical activity (PA) trackers and smartwatches can be used for health data collection in research as an addition to existing methods [1], and the data collected could be used to support patient diagnostics and treatment [2], see overview by Henriksen et al. [3].

For collecting data from many different device suppliers, a flexible and robust solution is needed, that can also receive data from heterogenous sources. The mSpider (Motivating continuous Sharing of Physical activity using non-Intrusive Data Extraction methods Retro- and prospectively) system is an experimental tool designed for automatic and continuous collecting of health-related data recorded by consumer-based activity trackers [4]. It has been designed to collect data of various PA-variables from activity trackers from a range of different providers. Today's activity trackers are smart devices capable of collecting many PA-variable estimates and transferring them to a smartphone for persistent storage. In their study, Henriksen et al. [4] collected smartwatch data using the mSpider system.

The current mSpider architecture consists of two servers, an administrative user-facing system (front-end), and a back-end server for gathering data by using the

[1] Corresponding Author: Erlend Johannessen, Department of Computer Science, UiT The Arctic University of Norway, PO Box 6050 Langnes, N-9037 Tromsø, Norway. E-mail: erlend.johannessen@uit.no.

manufacturers' public APIs. In addition, a mobile application has been made for those manufacturers (notably Apple and Samsung) where data only are available through SDKs provided by the manufacturers. **Figure 1** shows the original mSpider architecture.

In mSpider, the *participants* are enrolled in a *study*, after which they only need to wear their smart watches to collect and share data. Their activity data are uploaded to the device manufacturers' respective clouds and then pulled to the mSpider back-end server through the manufacturers' APIs.

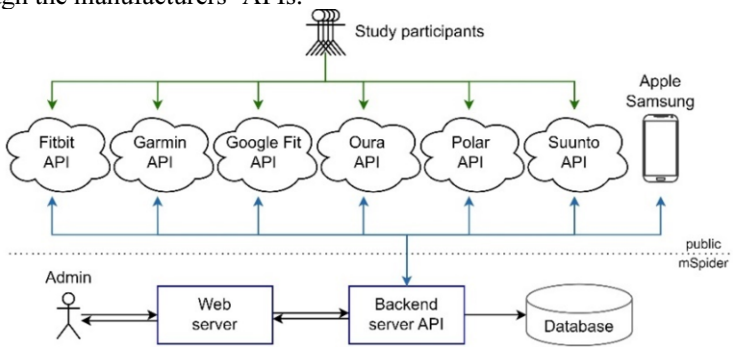


Figure 1. Original mSpider data collection architecture.

The programming language used for the mSpider back-end server was Go (go.dev, open source). The admin front-end used Node.js (nodejs.org, open source) for serving HTML, CSS, and JavaScript, and Angular (angular.io, open source) for creating the user interface. All data were stored in a MongoDB (mongodb.com, US) database. Both servers were run in a Docker (docker.com, US) environment using Docker-Compose on a single Ubuntu Linux (canonical.com, UK) server.

The goal of this study is to examine the mSpider system, identify problems and possible improvements, and to discuss an architecture capable of collecting data from a large population, over an extended period of time.

A digital twin is defined as a digital representation of physical entity. A *human* digital twin is maybe not fully realisable at the moment, but there is potential for creating small-scale human digital twins today, by joining different types of data from digital services and sources. The ambition aim is to add more data types and more data sources to the mSpider system, in order to approach a human digital twin system [5,6].

2. Methods

To analyse the state of the mSpider system and uncover problems, several experts were involved in interviews, and using the "think aloud" method, in a qualitative study approach [7], including the following steps:

1. Questioning the researchers using mSpider on how they experienced the system, and what they disliked or missed.
2. Discussing with the original developers and operational staff members behind mSpider, outlining the decisions governing the development of the system.
3. Reviewing the source code of the system, to understand the functionalities and state of the system.

3. Results

Observation done via researchers, developers (interview and code review), and operational staff are shown in **Table 1**. Priorities indicate the importance of the respective improvement. Issues are assigned to one type of actor, although some issues concerned or had consequences for several or all actors. Issues given low priority are not described in detail.

Table 1. Findings from researchers, developers, and operational staff, including code review. The rows are coloured differently to separate actors' observations from each other.

Actors	Observations	Priority
Operational staff	Security patches would be difficult to apply because of the way mSpider was developed and maintained	High
Operational staff	No assessment of threats or risks done on the system	High
Operational staff	Running system inside university network could theoretically put university network at risk	High
Operational staff	Every production update means deploying the full system. If something malfunctions the whole system may malfunction	Medium
Developers	Third-party components used got outdated and were no longer maintained and could not easily be replaced	High
Developers	Back-end server was a tightly coupled monolithic architecture	High
Developers	Back-end server was responsible for everything: participant consent dialogue, data collection, management, data extraction, batch runner for historical data	High
Developers	Device data were saved in the same storage, giving a format that did not suit all devices	Medium
Developers	Changing provider behaviour creates a new deployment of the full system	Low
Developers	Non-relational database may not be ideal for storage technology when there was need for combining several collections	Low
Developers	Different metadata were saved in the same storage collection	Low
Developers	Adding a new provider, initiated changes across the whole code base	Low
Researchers	A limited number of variables collected, e.g., daily step count, energy expenditure (kcal), and moderate or vigorous physical activity (PA)	Low / Medium
Researchers	Inefficient and cumbersome user interface	Low
Researchers	Only rudimentary data extraction from the system	Low
Researchers	Limited management functionality for study data	Low
Researchers	Limited management functionality for participant data	Low
Researchers	Data collection complexities with regards to when devices add their data to the manufacturer's cloud	Low

Several issues were uncovered from code reviews and from talking to developers. Device data from different manufacturers were saved into the same document storage, giving a general format that did not suit different data from different manufacturers. This required comprehensive mapping methods when reading from the different collections, since the various manufacturers have different data models. When used with a proof-of-concept system with limited data collection this worked but would be too complex when expanding on more data variables. Another issue was the use of third-party components in the source code. Using community code in your project is normally not an issue, but problems arise when packages get outdated and are no

longer maintained. This could be due to lack of security fixes, but also because the package is outdated with regards to functionality as to what the package was meant to solve. As an example, the Golang package used for MongoDB database access did not work with newer (and more secure) versions of MongoDB. This package was deeply integrated with the code and could not be easily replaced.

The back-end server (see **Figure 1**) was implemented as a monolith, which normally is not a problem, but a tightly coupled monolithic system tends to end up as a "big ball of mud" [8]. The server was responsible for everything, including data collection from APIs, being a receiving API for the mobile mSpider clients (for Apple and Samsung), being a management and data extraction system for researchers and admin personnel, and a batch runner for gathering historical data from the device providers. This is a lot of responsibility for a system and is difficult to maintain.

Operational staff were mainly concerned with security and deployment of the mSpider system, and there were some problems with the solution running inside the university network. Every production update meant deploying the full system. Theoretically, if something malfunctioned the whole system could malfunction. Security patches would be difficult to apply because of the way mSpider was developed and maintained. From reviewing the mSpider project we found that several security measures were implemented in mSpider, but there was no assessment of threats or risks. Because of this, data was collected and stored anonymously. Running the current system on a single Linux server on the university premises was a problem with regards to security for mSpider but could also pose a problem for the university's security.

Based on the expert-interviews and the source code review, we identified several requirements for a new system version. The new system should:

1) Be ready for productive use in a professional health context.
2) Be scalable with regard to new devices.
3) Be scalable with regard to data volume.
4) Collect more diverse data or groups of data.
5) Make data interpretable and easily available to researchers from various disciplines.

4. Discussion

Risk assessments are used to expose undesirable incidents in systems and evaluate probability of occurrence. To be production-ready the new system needs a thorough risk assessment, and the security needs to improve for the system not to risk leaking collected data. The system also needs to be running continuously, so a stable set of services is necessary. This is dependent on how the software architecture is implemented, and some software principles are essential for this, among them separation of concerns and extensive use of interfaces when using the relevant programming languages.

Using a secure and capable runtime environment is key, and a suggestion for this is running the solution in Microsoft Azure or another cloud computing service.

An important feature in the new system would be scalability with regards to new devices. One way of solving this would be to create a service for each data provider so that change to one device provider's API, access method, or data collected only would affect a single service. This would also make it easy to add new providers to the system,

in that the new provider could be developed and tested in isolation from the other services. Several services could also be added for each provider, so that the solution is scalable with regards to gathering increasing amounts of data from the population.

The storage system will be created such that data from each provider could be stored in their own databases. Creating the possibility for a separate database or localization of storage for each provider would increase scalability for storage, which again opens up for federated database technology [9].

Data collected are intended to be heterogeneous and the system should store as diverse and plentiful data as practically possible. The next version of mSpider should be expanded so that it covers more data types (e.g., pulse, sleep, temperature, body composition) and makes a low-resolution human digital twin [5] possible.

5. Conclusion

We have identified requirements for a production ready, scalable, and flexible data collection system for personal devices. One of the overarching goals for the new mSpider system is to enable it for population-based research investigating potential changes in a population's lifestyle and health. This would imply continuous data collection from the population to create data-driven analyses, working towards a human digital twin [6]. Researchers should be able to create a research project by initially setting some parameters for what data they want to be included in the study, such as steps, heart rate, weight, and sleep. These data could be extracted daily into a warehousing system [10]. This data collection system could give new opportunities for public health research.

References

[1] Brickwood K-J, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. JMIR MHealth UHealth. 2019;7:e11819. doi: 10.2196/11819.

[2] Gwaltney C, Coons SJ, O'Donohoe P, O'Gorman H, Denomey M, Howry C, Ross J, ePRO Consortium. "Bring Your Own Device" (BYOD): The Future of Field-Based Patient-Reported Outcome Data Collection in Clinical Trials? Ther Innov Regul Sci. 2015;49:783–791. doi: 10.1177/2168479015609104.

[3] Henriksen A, Mikalsen M, Woldaregay A, Muzny M, Hartvigsen G, Hopstock L, Grimsgaard S. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. J Med Internet Res. 2018;20:e110. doi: 10.2196/jmir.9157.

[4] Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA. Consumer-Based Activity Trackers as a Tool for Physical Activity Monitoring in Epidemiological Studies During the COVID-19 Pandemic: Development and Usability Study. JMIR Public Health Surveill. 2021;7:e23806. doi: 10.2196/23806.

[5] Johannessen E, Henriksen A, Hartvigsen G, Horsch A, Årsand E, Johansson J. Ubiquitous digital health-related data: clarification of concepts. Scand Conf Health Inform. 2022;52–58. doi: 10.3384/ecp187009.

[6] Miller ME, Spatz E. A unified view of a human digital twin. Hum-Intell Syst Integr. 2022;4:23–33. doi: 10.1007/s42454-022-00041-x.

[7] Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qual Res Sport Exerc Health. 2017;9:514–531. doi: 10.1080/2159676X.2017.1331501.

[8] Big Ball of Mud [Internet]. [cited 2022 Nov 14]. Available from: http://www.laputan.org/mud/.

[9] Sheth AP, Larson JA. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Comput Surv. 1990;22:183–236. doi: 10.1145/96602.96604.

[10] Chaudhuri S, Dayal U. An overview of data warehousing and OLAP technology. ACM SIGMOD Rec. 1997;26:65–74. doi: 10.1145/248603.248616.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230287

Standardization Proposal for the Transmission of Waiting List Data in Italy

Ylenia MURGIA[a], Monica BONETTO[b], Roberta GAZZARATA[b,c], Luisa BROGONZOLI[d], Rosaria IARDINO[d], Alessandro VENTURI[e,f] and Mauro GIACOMINI[a,b,1]

[a] *Dep. Of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy*
[b] *Healthropy s.r.l., Savona, Italy*
[c] *HL7 Europe, Brussels, Belgium*
[d] *The Bridge Foundation, Milan, Italy*
[e] *Dep. of Political and Social Sciences, University of Pavia, Pavia, Italy*
[f] *I.R.C.C.S. Policlinico San Matteo, Pavia, Italy*

Abstract. Each Italian region is required to manage and disclose data relating to waiting times for healthcare services which are provided by both public and private hospitals and local health units accredited to the *Sistema Sanitario Nazionale* (SSN – in English, National Healthcare System). The current law governing data relating to waiting times and their sharing is the *Piano Nazionale di Governo delle Liste di Attesa* (PNGLA – in English National Government Plan for Waiting Lists). However, this plan does not propose a standard to monitor such data, but only provides a few guidelines that the Italian regions are required to follow. The lack of a specific technical standard for managing sharing of waiting list data and the lack of precise and binding information in the PNGLA make the management and transmission of such data problematic, reducing the interoperability necessary to have an effective and efficient monitoring of the phenomenon. The proposal for a new standard for the transmission of waiting list data derives from these shortcomings. This proposed standard promotes greater interoperability, is easy to create with an implementation guide, and has sufficient degrees of freedom to assist the document author.

Keywords. Waiting list data, interoperability, standard

1. Introduction

Monitoring waiting list data in healthcare has great relevance in ensuring all citizens have access to health services that are insured and adequate for the clinical problems presented. One of the main tasks of the *Servizio Sanitario Nazionale (SSN* – in English, National Healthcare System*)* concerns with meeting citizens' healthcare needs and providing transparency regarding waiting times. As a matter of fact, public and private bodies, that provide health care on behalf of the SSN, are required to show on their website both the expected waiting times and the actual average waiting times for each type of service provided [1]. Currently, the waiting lists and waiting times issues are governed by the

[1] Corresponding Author, Mauro Giacomini, DIBRIS, University of Genoa, Via all'Opera Pia 13, 16145 Genoa, Italy; E-mail: mauro.giacomini@dibris.unige.it.

Piano Nazionale di Governo delle Liste di Attesa (PNGLA – in English National Government Plan for Waiting Lists*)*, which is a plan designed to guarantee citizens fair, appropriate, and timely access to health services limiting access times [2], even though often last longer than necessary [3]. The current edition of the plan straddles the pre-COVID-19 and post-COVID-19 period; therefore, if under ordinary conditions waiting times may not be respected, during the COVID-19 emergency this problem is even more exasperated and both waiting times and waiting lists have been further extended. For example, the Organisation for Economic Co-operation and Development (OECD), which offers extensive data on this topic, states that during the pandemic, patients' time on waiting lists for several services increased [4]. Moreover, the OECD recognizes that it is difficult to find common definitions of criteria to detect waiting lists [5]. Indeed, waiting list data's major problem is that waitlist statistics not only remain fragmented and difficult to compare across countries, but in Italy, they even are incomparable among different regions. Italian healthcare is organized at a regional level, and in accordance with the law, regions should publish waiting list data to make healthcare public administration data, documents, and information available [6], but this requirement is not always satisfied. In February 2021, The Bridge Foundation started a project focused on the 2019 and 2020 waiting list data collection and comparison [7]. During this two-year analysis, heterogeneity was found in the data provided by the regions, and moreover, a change in the method of collecting information from one year to the next was also observed. The current version of PNGLA obliges NHS-accredited health units and hospitals that provide health services to disclose information regarding waiting lists, but this plan contains too many degrees of freedom: there are not enough indications to ensure that all regions can present homogenous datasets. Therefore, this work does not aim to find a solution to long waiting times in healthcare but to improve the management and sharing of waiting lists.

2. Materials and Methods

To allow the development of an efficient and effective tool for monitoring and sharing data regarding waiting times in healthcare, particular attention was paid to the PNGLA. The plan identifies a list of healthcare services, that are subject to controls on waiting times, divided into outpatient services and hospitalization services and, in turn, outpatient services are divided into specialist visits and instrumental performances. The plan also establishes the maximum times within which they must necessarily be provided, depending on the priority assigned at prescription time. Therefore, patients considered more serious are subject to shorter waiting times, compared to less urgent cases. In the specific case of outpatient services, priorities U, B, D, and P are used (in decreasing order of urgency), while for hospitalizations, classes A, B, C, and D are used (in decreasing order of urgency). Considering these guidelines, a standardization method for the management and sharing of waiting list data is therefore proposed to foster interoperability, a mandatory requirement stated in Italian law [8]. Finding a proper document structure for the creation of the new standard was not easy. An Italian law of 2015 [9] requires the adoption of HL7 Clinical Document Architecture (CDA) for all documents to be recorded in the Regional Health Informatic Infrastructures. Therefore, the authors decided to take a cue from the CDA to design a new draft standard document also following the suggestion of some colleagues of *Liguria Digitale* (the informatic public company of Liguria Region). These colleagues also provided us with raw data to

be able to carry out tests, , specifically designed to monitor and share data regarding waiting times for healthcare services. As the CDA is an Extensible Markup Language (XML) based markup standard, we decided to create a new document based on XML. To create the structure of the document we used the Integrated Development Environment (IDE) Visual Studio Code. After implementing the general structure of the XML document, containing everything necessary for the correct monitoring of data regarding waiting times in healthcare, we created an XML Schema Definition (XSD) to validate the document, again using Visual Studio Code as IDE. Since the raw data provided by *Liguria Digitale* were in XLSX format, we designed a program that allows the conversion of this file into a document in XML format. To build this program, we used Python Programming Language, using PyCharm as IDE.

3. Results

Therefore, for the sharing of waiting list data, we created a new draft standard, the Waitlist Document Architecture (WDA). Taking a cue from the CDA, also the WDA document consists of two major sections: the WDA header and the WDA body. The header identifies the document type and provides information on the figures involved, such as the region (Figure 1), the author and the custodian of the document, and the period (Figure 2), which contains information about the year and the month or quarter to which the collected data refers. To uniquely identify the figures involved, specific, already existing codes must be used. For example, to identify the regions, the statistical code of territorial administrative units (provided by the Italian Institute of Statistics - ISTAT) must be used, while to identify the author and the custodian of the document, the tax code must be used.

```
<recordTarget>
    <regionRole>
        <id code="070" codeSystemName="ISTAT" assigningAuthorityName="Liguria"/>
        <region>
            <name>Liguria</name>
        </region>
    </regionRole>
</recordTarget>
```

Figure 1. Region coding example.

```
<period>
    <year>2020</year>
    <month>02</month>
</period>
```

Figure 2. Period coding example.

The body is a structured XML body, which contains information regarding the healthcare services, relative values on waiting times, and the health facility that provided these services. We designed two different WDA documents, according to the type of monitoring. For outpatient services, it is possible to insert two specialty sections, one concerning specialist visits and the other concerning instrumental performance. Instead,

for hospitalizations, it is possible to insert a single specialty section, concerning only hospitalization services. We decided for this option to consider the already existing different nomenclatures of healthcare services. Indeed, specialist visits and instrumental services are coded in compliance with *Livelli Essenziali di Assistenza (LEA* – in English Essential Assistance Levels) codes, while hospitalizations are identified according to the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) codes (Figure 3).

```
<component>
    <section code="89.7A.3" codeSystemName="LEA" displayName="Prima Visita cardiologica">
        <title>PRIMA VISITA CARDIOLOGICA</title>
```

Figure 3. Example of use of the terminology of healthcare services.

Accordingly, each specialty section must contain several "leaf" sections, containing the name of the service and the corresponding LEA code or ICD-9-CM code. For each region, each healthcare service can be provided by several health institutions or facilities. Therefore, within each "leaf" section there is a CLUSTER-type organizer tag, which can be replicated for each type of structure and/or facility present, and each facility or structure is associated with a unique identification code. Both the monitoring of waiting times for outpatient services and for hospitalization services require the inclusion of the priority class of the service in question. Therefore, for each structure or facility, it is possible to enter several observations of the same service associated with a specific priority class (Figure 4). Although the use of priority classes is mandatory, it often happens that it is not specified. For this reason, we decided to insert a generic observation, which does not specify the type of priority of the healthcare service in question. If the priorities associated with the service in question are managed, the observations must be completed with one of the codes of the HL7 ActPriority dictionary, characterized by its own Object Identifier (OID).

```
<priorityCode code="A" codeSystem="2.16.840.1.113883.5.7"
codeSystemName="Act Priority" displayName="ASAP">
    <translation code="B" codeSystem="2.16.840.1.113883.2.9.5.2.3"
    codeSystemName="Priorità Visita" displayName="Breve"/>
</priorityCode>
```

Figure 4. Example of using priority classes.

Finally, for each observation, there must be at least one of the number of reservations guaranteed on schedule according to the type of priority (if any), or the ratio between the number of reservations guaranteed on schedule, compared to the total number of reservations (guarantee percentage). Optionally, the total number of reservations in the month or in the quarter in question can be added (Figure 5).

```
<values>
    <totalNumber>143</totalNumber>
    <mean>136</mean>
    <percentage>95,10</percentage>
</values>
```

Figure 5. Example observation values.

Then, the validation of the WDA document makes it possible to verify if all the mandatory information is present, and if the data entered is correct. Furthermore, to help the author of the document write a WDA document, we created an Implementation Guide, containing the instructions necessary to write both the header and the body of the document.

4. Discussion and Conclusion

This project is currently still under development to try to make further improvements to facilitate the use of this standard. However, simulations have already been carried out with already-acquired data, provided by *Liguria Digitale*. Converting the data into XML format, allows us to implement more in-depth checks on the quality of the data itself, allowing us to have information that is not only available but also usable. This proves the validity of the project and that it is legitimate to assume that we can continue with further analysis and investigations. This proposal is based on the present PNGLA relating to 2019-2021, since at the time of writing this article, the version of the plan for the 2022-2024 period is not yet available. Therefore, for the next few years, the authors expect greater standardization also on the part of guidelines to ensure greater interoperability and constant monitoring and transmission of data. A further aim is to propose the WDA document as a new standard for transmitting waiting list data to HL7 Italy to create a turning point in the future of this topic. At present, this document is intended to represent only the structure of a possible new data-sharing standard but implementing a FHIR profile with APIs in the future could be a significant evolution in terms of easier implementation. Another possible evolution could be also the adoption of a terminology management system to consider future versions of applied terminology [10]. Moreover, we are looking for enthusiastic young people who are willing to learn how to implement our standard and bring innovative ideas that can help the development of this proposal.

References

[1] Decreto Legislativo 14 marzo 2013, n. 33, in: art. 41, c. 6, Italian Republic.
[2] Piano Nazionale di Governo delle Liste di Attesa: allegato A [Internet]. Roma: Ministero della Salute; 2019. Available from: http://www.salute.gov.it/imgs/C_17_pubblicazioni_2871_allegato.pdf
[3] Fattore G, Mariotti G, Rebba V. Review of waiting times policies. Country case studies: Italy. In: Siciliani L, Browitz M, Moran V, eds. Waiting Time Policies in the Health Sector: What Works? OECD Publishing; 2013. p. 167-181.
[4] OECD. Health at a Glance 2021: OECD indicators [Internet]. Paris: OECD Publishing; 2021. Available from: https://doi.org/10.1787/health_glance-2021-en
[5] Hurst J, Siciliani L. Tackling Excessive Waiting Times for Elective Surgery: A Comparison of Policies in Twelve OECD Countries. Paris: OECD Publishing; 2003.
[6] Decreto Legislativo 14 marzo 2013, n.33, in: art.5, Italian Republic.
[7] Bonetto M, Maggi N, Fuschi D, Venturi A, Brogonzoli L, Iardino R, Giacomini M. Healthcare Insights:
Evaluating the Access to the Italian Healthcare System. In: Studies in Health Technology and Informatics. IOS Press BV; 2022. p. 709–10.
[8] Decreto Legislativo 14 marzo 2013, n. 33, in: art. 7, Italian Republic.
[9] Decreto Del Presidente Del Consiglio Dei Ministri 12 febbraio 2015, n.178, Italian Republic.
[10] Gazzarata R, Eugenia Monteverde M, Vio E, Gubian L, Borgo I, Giacomini M. A Terminology Service Compliant to CTS2 to Manage Semantics within the Regional HIE. Vol. 13, EJBI. 2017.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230288

OpenDataPsy: An Open-Data Repository with Standardized Storage and Description for Research in Psychiatry

Chloé SAINT-DIZIER[a,b], Antoine LAMER[a,b,c], Majda ZAANOUAR[b,c], Alina AMARIEI[a] and Paul QUINDROIT[c,1]
[a] *Fédération régionale de recherche en psychiatrie et santé mentale (F2RSM Psy), Hauts-de-France, Saint-André-Lez-Lille, France*
[b] *Univ. Lille, Faculté Ingénierie et Management de la Santé, F-59000, Lille, France*
[c] *Univ. Lille, CHU Lille, ULR 2694, METRICS: Évaluation des Technologies de santé et des Pratiques médicales, F-59000, Lille, France*

Abstract. Sharing health data could avoid duplication of effort in data collection, reduce unnecessary costs in future studies, and encourage collaboration and data flow within the scientific community. Several repositories from national institutions or research teams have making their datasets available. These data are mainly aggregated at spatial or temporal level, or dedicated to a specific field. The objective of this work is to propose a standardized storage and description of open datasets for research purposes. For this, we selected 8 publicly accessible datasets, covering the fields of demographics, employment, education and psychiatry. Then, we studied the format, nomenclature (i.e., files and variables names, modalities of recurrent qualitative variables) and descriptions of these datasets and we proposed on common and standardized format and description. We made available these datasets in an open gitlab repository. For each dataset, we proposed the raw data file in its original format, the cleaned data file in csv format, the variables description, the data management script and the descriptive statistics. Statistics are generated according to the type of variables previously documented. After one year of use, we will evaluate with the users if the standardization of the data sets is relevant and how they use the dataset in real life.

Keywords. Open data, psychiatry, data reuse, research

1. Introduction

Sharing health data could avoid duplication of effort in data collection, reduce unnecessary costs in future studies, and encourage collaboration and data flow within the scientific community. It would also improve the reproducibility and transparency of clinical research by allowing researchers to validate each other's results and reduce the impact of publication bias [1]. Many institutions are now making their datasets available. For example, in France, open platform contains French public data about demographics, COVID-19, elections, energy consumption, health and employment [2]. Researchers are increasingly encouraged to deposit their work on these platforms. These data were previously crossed for a study on the relationship between the immigrant rate and health

[1] Corresponding Author: Paul QUINDROIT, E-mail: paul.quindroit@univ-lille.fr.

status [3]. Institutions data are generally aggregated at a geographical level (e.g., a city or region) over a defined period (e.g., day, month, year). More specific repositories also propose individual data [4], from generic domain or specific data as hospital intensive care units [5–7]. The datasets are spread over different repositories, with heterogeneous formats, nomenclatures and descriptions.

The objective of this work is to propose a standardized storage and description of open datasets dedicated to psychiatry and mental health.

2. Methods

We have selected useful datasets for research in psychiatry available as open access on various national or international data sharing platform. They provide ecological variables which often supplement individual psychiatric variables (e.g., in a cohort) to gain information about their environment and adjust the statistical models. We studied the format, nomenclature (i.e., files and variables names, modalities of recurrent qualitative variables) and descriptions of these datasets.

We had to select a unique file format (i.e., type of file, column separator, decimal character, header, encoding), standardized files names, variables names and description. We also provided a standardized description of the variables composing the data file with a non ambiguous and meaningful variable name for printing and the type of variables (i.e., binary, qualitative, continuous quantitative, discrete quantitative). We did not take into account non relevant variables as unique identifier, qualitative variable with too much modalities and variables (e.g., zipcodes) corresponding to the statistical unit of the file. For each of the datasets, we kept the original dataset and we added the standardized file. We also provided a data management script when necessary. We developed a script to automatically generate descriptive statistics based on the cleaned and managed data file and the description of the variables. Statistics are generated according to the type of variables previously documented.

All the documents are available in a gitlab directory, with one folder per dataset and a README file to describe each folder [8].

3. Results

3.1. Source files

We selected 8 datasets, covering the fields of demographics, employment, education and psychiatry. We also provided a dataset giving a correspondence between two French zip codes nomenclatures. Raw files were available in heterogeneous formats (csv in French or International format, xls, xlsx), delimiting character in the case of flat files, decimal character (“,” and “.”) and encoding (LATIN-1, LATIN-3 and UTF8). Some datasets formats were often of poor quality, with empty lines at the top or bottom, specifying the source and the generation date. A file presenting social disadvantage indicator was generated based on several other variables: tax income, unemployment, workers, graduates.

3.2. Standardization

Each repository contains 5 files: the raw data file in its original format with a prefix for the type of file and a suffix for the years covered by the file, the cleaned data file in csv format, the variables description, the data management script and the descriptive statistics. The table 1 provides the standardized names and description of each file.

Standardization of the cleaned data file complies with the following characteristics:

- File format: csv, column separator ",", decimal character "."
- Variable names: lowercase, no special characters except the underscore
- Standard modalities for sex (M/F)
- Date in YYYY-MM-DD (year, month, day)
- Frequent Standardization of variable names for the most frequent ones (e.g., insee_code, postal_code).

Each dataset was documented with a README file at the root of the folder. The elements of this file are described in table 2, The figure 1 represents the folder dedicated to the finess *(FIchier National des Etablissements Sanitaires et Sociaux, French designation for the national register of health and social establishments)* dataset, a list of all French health facilities. The figure 2 is the subdirectory containing the raw finess dataset and the R data management file.

Table 1. Standardized name and description of files available for each dataset

Files	Description
raw_XXX_year.xslx	Source file without modification
dm_XXX_year.R	Data management script for cleaning and standardizing the raw data file
cleaned XXX_year.csv	Cleaned and standardized file
variables_XXX_year.csv	Description of the variables
desc_XXX_year.html	Descriptive statistics

Table 2. Documentation items contained README file

Section	Description
Description	Context (e.g., country and date of generation), productor (e.g., institution, company), statistical unit, type of data
Overview	Display of the five first rows
License and reuse conditions	Name of the license and/or link with the source page with use conditions
Raw data source	Link to the source page where the data were downloaded
References	Link to data productor website or bibliographic references

Table 3. Statistical indicators and graphics appropriate for each type of variable

Type of variable	Statistics	Graphics
Binary variable	Count, percentage, percentage of missing data	Doughnut
Qualitative variable	Count, percentage, percentage of missing data	Barplot
Continuous quantitative	Median IDR, min, max, percentage of missing data	Histogram and density
Discrete quantitative	Median IDR, min, max, percentage of missing data	Barplot
Date	Min, max, number of events per date with quartile	-

3.3. Automated descriptive statistics

For each type of variable, we propose appropriate statistical indicators and graphics, described in table 3. The figure 3 represents descriptive statistics for the variable "Number of bed-days for full time hospitalization", available in the dataset dedicated to annual hospital activity in psychiatry.

Name	Last commit	Last update
..		
raw	Clean R scripts	2 weeks ago
README.md	Update readme	4 weeks ago
cleaned_finess_2021.csv	Rename cleaned files	2 weeks ago
stat_desc_cleaned_finess_2021_.html	Scripts descriptive stats	1 week ago
variables_finess_2021.csv	Rename cleaned files	2 weeks ago

Figure 1. Repository for finess dataset, a list of all French health facilities.

Name	Last commit	Last update
..		
dm_finess_2021.R	Clean R scripts	2 weeks ago
head_finess_2021.png	Update readme	1 month ago
raw_finess_2021.csv	Add "finess" directory	1 month ago

Figure 2. Subdirectory containing the raw finess dataset and the R data management file.

```
## #  Available bed-days for full time hospitalisation
##
##  type : quantitative continue
##
## Missing data :  17.56098 %
## Quartile :
```

min	Q1	median	Q3	max
0	15330	28470	48460.5	347181

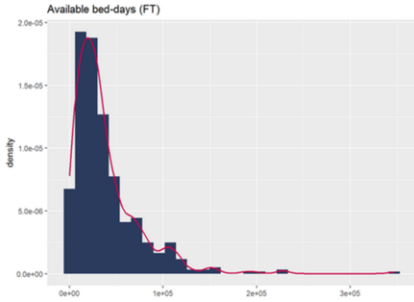


Figure 3. Automated descriptive statistics for the dataset dedicated to annual hospital activity.

4. Discussion and Conclusions

In this study, we implemented OpenDataPsy, a standardized storage and description of open datasets dedicated to psychiatry and mental health. These datasets are directly ready to be used, after being transformed into a common standard format. The description of

variables allowed to automatically generate descriptive statistics of each dataset, in common format.

It remains to manage the recovery of the data sets when they are updated on the initial source site. We also need to spread the repository inside the psychiatric community. After several months of use, we will evaluate with the users if the standardization of the data sets is relevant and how they use the dataset in real life. In the future, we will feed the directory with new datasets. In particular, we plan to integrate free text data coming from forum and social media. It will be necessary to anonymise this data before sharing.

References

[1] Shahin MH, Bhattacharya S, Silva D, Kim S, Burton J, Podichetty J, Romero K, Conrado DJ. Open data revolution in clinical research: opportunities and challenges. Clinical and Translational Science. 2020 Jul;13(4):665-74. doi:10.1111/cts.12756.

[2] Inventaires thématiques de données - data.gouv.fr, (n.d.). https://www.data.gouv.fr/fr/pages/thematiques-a-la-une/ (accessed November 17, 2022).

[3] Perrot J, Hamel JF, Lamer A, Levaillant M. The Relationship between the Immigrant Rate and Health Status in the General Population in France. Journal of Personalized Medicine. 2021 Jun 30;11(7):627. doi:10.3390/jpm11070627.

[4] Banzi R, Canham S, Kuchinke W, Krleza-Jeric K, Demotes-Mainard J, Ohmann C. Evaluation of repositories for sharing individual-participant data from clinical studies. Trials. 2019 Dec;20:1-0. doi:10.1186/s13063-019-3253-3.

[5] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation. 2000 Jun 13;101(23):e215-20.. doi:10.1161/01.cir.101.23.e215.

[6] Johnson A, Bulgarelli L, Pollard T, Celi LA, Mark R, Horng IV S. MIMIC-IV-ED. PhysioNet. 2021. doi:10.13026/S6N6-XD98.

[7] Thoral PJ, Peppink JM, Driessen RH, Sijbrands EJ, Kompanje EJ, Kaplan L, Bailey H, Kesecioglu J, Cecconi M, Churpek M, Clermont G. Amsterdam University Medical Centers Database (AmsterdamUMCdb) Collaborators and the SCCM/ESICM Joint Data Science Task Force: Sharing ICU patient data responsibly under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example. Crit Care Med. 2021 Jun 1;49(6):e563-77. doi:10.1097/CCM.0000000000004916.

[8] health_data_science / open_data_psy · GitLab, GitLab. (n.d.). https://gitlab.com/d8096/open_data_psy (accessed January 4, 2023).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230289

Description of a French Population of Diabetics Treated Followed up by General Practitioners

Mathilde FRUCHART[a], Antoine LAMER[a] , Madleen LEMAITRE[a,b], Jean-Baptiste BEUSCART[a], Matthieu CALAFIORE[a,c] and Paul QUINDROIT[a1]
[a] *Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France*
[b] *CHU Lille, Department of Diabetology, Endocrinology, Metabolism and Nutrition Lille University Hospital, F-59000 Lille, France*
[c] *Department of General Practice, University of Lille, Lille, F-59000 Lille, France*

Abstract. In France, the prevalence of treated diabetes has been estimated at 4.6%, or more than 3 million people and 5.2% in Northern France. The reuse of primary care data allows to study outpatient clinical data such as laboratory results and drug prescriptions, which are not documented in claims and hospital databases. In this study, we selected the population of treated diabetics from the Wattrelos primary care data warehouse, in North of France. Firstly, we studied the laboratory results of diabetics by identifying whether the recommendations of the French National Authority for Health (HAS) were respected. In a second step, we studied the prescriptions of diabetics by identifying the oral hypoglycemic agents treatments and insulins treatments. The diabetic population represents 690 patients of the health care center. The recommendations on labortatory are respected for 84% of diabetics. The majority of diabetics are treated with oral hypoglycemic agents 68.6%. As recommended by the HAS, metformin is the first-line treatment in the diabetic population.

Keywords. Diabetes mellitus, primary care, general practitioner, data reuse

1. Introduction

In France, the prevalence of treated diabetes has been estimated at 4.6%, or more than 3 million people and 5.2% in Northern France [1]. This prevalence increases with age and, at equal age, is higher in men than in women. Over the last 10 years, the number of people treated for diabetes has increased by an average of 5%/year. Within the management and follow-up of diabetes, blood and urine tests are necessary. The French National Authority for Health (HAS) recommends at least twice a year the measurement of glycated hemoglobin and once a year a renal evaluation (measurement of creatinine) [2].

For about ten years, health data warehouses have been implemented in hospitals and contain all the data collected during the stay (laboratory results, biometric measurements, physicians' letters, diagnosis). However, the data is specific to the hospital and no follow-

[1] Corresponding Author: Quindroit Paul, E-mail: paul.quindroit@univ-lille.fr.

up outside the hospital is recorded. In contrast, data warehouses based on claims databases contain data on the reimbursement of care for all health insurance beneficiaries but do not contain clinical data [3]. Primary care data refers to all clinical data outside the hospital setting and provides an overview of patient follow-up over several years. It includes data from a consultation with a General Practitioner (GP), prescriptions for drug and laboratory results. The reuse of primary care data is increasingly being addressed, but only a few retrospective studies were conducted with this kind of data [4,5]. The French project PriCaDa aims to reuse primary care data to improve GP practices in collaboration with the multidisciplinary health centre (MHC) in Wattrelos, in North of France. The objective of this study is to use the data warehouse to analyze data from the diabetic population of the Wattrelos MHC.

2. Methods

We implemented a data warehouse based on the Wattrelos MHC [6]. It contains data from the patients' medical records of six GPs (year of birth, sex, reason for consultation and data from the consultation, biometric measurements taken at the consultation as weight/height, drug prescriptions, laboratory results, medical history, letters prescribing external medical procedures). An observation phase with GPs was carried out to understand data capture and to identify meaningful and usable variables.

The data were transformed and loaded into the common data model Observational Medical Outcomes Partnership (OMOP) [7]. The source data contained a vocabulary specific to the health care structure, filled in free text or in a local terminology. These anonymised data are stored on a secure server. The collection, transformation and loading of this data is managed automatically by an extract-transform-load (ETL) process, implemented in Python and PostgreSQL.

The diabetic population is identifiable by different methods and is GP-dependent. For example, some GPs indicate the disease in the medical histories section with ICD10 codes or free text. Other GPs indicate the disease in the diagnosis consultation section in free text.

Our inclusion criteria were adult patients (age >= 18 years old) treated with anti-diabetics (oral hypoglycemic agents or insulin) over a 5-year period (from 2015 to 2020). We selected patients who had at least one prescription for drugs with an ATC code starting with A10, corresponding to the Drugs used in diabetes class [8]. We extracted prescribed drugs, lab test results, biometric measurements taken during the consultation of this population for the five years.

Firstly, we assessed if diabetics complied with the recommendations of the HAS on laboratory tests, in computing the number of patients with at least one creatinine result, the number of patients with more than two glycated haemoglobin (HbA1c) results and the number of patients with a HbA1c result per year [9]. In a second step we studied general prescriptions (all drugs) and antidiabetic prescriptions on the population of the year 2020. For the prescriptions, we focused on the year 2020 and we extracted prescriptions of oral hypoglycemic agents (OHAs; ATC codes started with A10 except A10A) and prescriptions of the insulin treatments (ATC codes started with A10A).

3. Results

3.1. Diabetic population

Over the period 2015 to 2020, 9,435 adult patients consulted in the Wattrelos MHC. The adult diabetic population represents 690 patients (7.3% of adult patients) treated with anti-diabetic drugs. The mean age (standard-deviation) of these patients is 65.7 (13.2) years old at the consultation and the population includes 46.8% women.

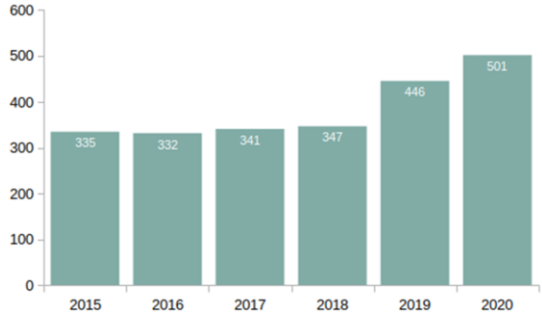


Figure 1. Number of diabetic patients per year

Diabetics population represents a mean (standard-deviation) of 3,152 (785.4) consultations per year (12.8% of the adult population), 15,596 (2,387.9) drugs prescriptions (23.5 % of the adult population) and 21,031 (2,699.8) laboratory results per year (18.5 % of the adult population). The Figure 1 represent the number of diabetic patients per year. Regarding the HAS recommendations, on average 84.4% of diabetic patients had more than two HbA1c laboratory results and 84.3% had at least one creatinine laboratory result.

Second paragraph.

3.2. Diabetes prescriptions for the year 2020

The majority of patients (68.6%) had only OHAs prescribed and 22. 2% had prescriptions for both insulin and OHAs during the year (Figure 2).

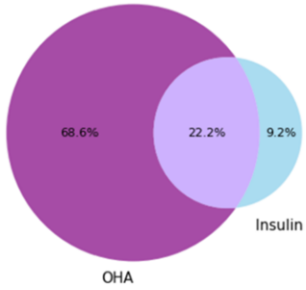


Figure 2. Distribution of patients by category of anti-diabetic prescriptions in 2020.

Concerning general prescriptions, (without anti-diabetics prescriptions), paracetamol (7.0%), acetylsalicylic acid (4.4%) and atorvastatin (3.5%) are the most prescribed (Figure 3). For the two categories of antidiabetics, metformin (43.8% of OHAs) and insulin glargine (41.5% of insulin treatments) were the most prescribed (Figure 3).

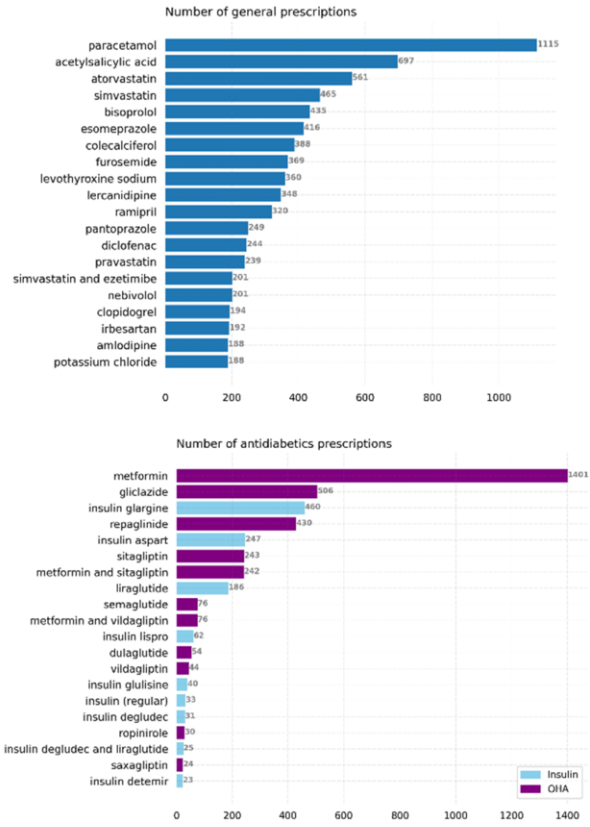


Figure 3. Distribution of the most prescribed drugs in 2020 (up: general prescriptions, down:by category of anti-diabetic prescriptions).

4. Discussion and Conclusions

We demonstrated the possibility to reuse the data of a primary care data warehouse to describe a diabetic population, which is currently rarely performed in France. The prevalence of diabetes in the Wattrelos MHC is higher than in the north of France (8.2% vs 5.2% in 2020). Moreover, diabetics account for a significant proportion of GP consultations, prescriptions and laboratory tests. The results also show that the HAS recommendations are relatively well followed. It should be noted that some laboratory tests are prescribed by medical doctors other than those of the MHC and that the results are sometimes not transmitted to the GPs. Finally, as recommended by the HAS, metformin is the first-line treatment in the diabetic population [2].

This study had several strengths. Firstly, we studied data in general practice with a large number of diabetic patients (n=690). This is the only source of data in which outpatient laboratory results may be queried. Secondly, we have a long period of data (5 years), that enabled us to study the follow-up of the number of consultations over time. Thirdly, the reuse of real-life data indicates that our results could be extended to outpatients more generally to assess compliance with the recommendations.

The present study had several limitations. Firstly, we are not able to distinguish between type 1 diabetes, type 2 diabetes or other types of diabetes. This information is usually filled in by free text by GPs or is missing from the data. Secondly, we considered only pharmacologically treated diabetics. For the same reason, diabetics under preventive hygiene and dietary measures are not documented in a structured format and are therefore difficult to reuse. For now, data came from a single health center but this study will be extended to other centres of the region.

Acknowledgements and Funding

The authors want to thank all the participants for their availability. This work was funded by the National Center for Precision Diabetic Medicine (PreciDIAB), which is funded jointly by the French Government (under the Investissements d'Avenir (Investments for the Future) program managed by the Agence Nationale de la Recherche (ANR, French National Research Agency; reference: ANR-18-IBHU-0001)), the European Union (through the European Regional Development Fund (ERDF; grant reference: NP0025517)), the Conseil Régional des Hauts-de-France (Hauts-de-France Regional Council; grant reference: 20001891/NP0025517), and the Métropole Européenne de Lille (MEL, European Metropolis of Lille; grant reference: 2019_ESR_11).

References

[1] Fuentes S, Mandereau-Bruno L, Regnault N, Bernillon P, Bonaldi C, Cosson E, et al. Is the type 2 diabetes epidemic plateauing in France? A nationwide population-based study. Diabetes & Metabolism. 2020 Nov 1;46(6):472–9.

[2] Haute Autorité de Santé. Prévention et dépistage du diabète de type 2 et des maladies liées au diabète [Internet]. Haute Autorité de Santé. 2015 [cited 2023 Jan 9]. Available from: https://www.has-sante.fr/jcms/c_2012494/fr/prevention-et-depistage-du-diabete-de-type-2-et-des-maladies-liees-au-diabete

[3] Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 2017 Aug;26(8):954–62.

[4] Boullenger L, Quindroit P, Legrand B, Balcaen T, Calafiore M, Rochoy M, et al. Type 2 diabetics followed up by family physicians: Treatment sequences and changes over time in weight and glycated hemoglobin. Prim Care Diabetes. 2022 Oct;16(5):670–6.

[5] Chazard E, Marcolino MS, Dumesnil C, Caron A, Palhares DM, Ficheur G, Marino BC, Alkmim MB, Beuscart R, Ribeiro AL. One million electrocardiograms of primary care patients: a descriptive analysis. InMedInfo 2015 Jan 1 (pp. 69-73).

[6] Fruchart M, Quindroit P, Patel H, Beuscart JB, Calafiore M, Lamer A. Implementation of a Data Warehouse in Primary Care: First Analyses with Elderly Patients. Stud Health Technol Inform. 2022 May 25;294:505–9.

[7] Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574–8.

[8] WHOCC - ATC/DDD Index, (n.d.). https://www.whocc.no/atc_ddd_index/?code=A10 (accessed January 10, 2023).

[9] French National Health Insurance Agency for Salaried Workers. Diabète : quelles analyses de sang et d'urines ? [Internet]. 2021 [cited 2023 Jan 10]. Available from: https://www.ameli.fr/lille-douai/assure/sante/themes/diabete-suivi/analyses-sang-urines

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230290

Digital Disease Surveillance for Emerging Infectious Diseases: An Early Warning System Using the Internet and Social Media Data for COVID-19 Forecasting in Canada

Yang YANG[a,1] Shu-Feng TSAO[a] Mohammad A. BASRI[b] Helen H. CHEN[a] and Zahid A. BUTT[a]
[a]*School of Public Health Sciences, University of Waterloo, Canada*
[b]*Systems Design Engineering, University of Waterloo, Canada*

Abstract. Background: Emerging Infectious Diseases (EID) are a significant threat to population health globally. We aimed to examine the relationship between internet search engine queries and social media data on COVID-19 and determine if they can predict COVID-19 cases in Canada. **Methods:** We analyzed Google Trends (GT) and Twitter data from 1/1/2020 to 3/31/2020 in Canada and used various signal-processing techniques to remove noise from the data. Data on COVID-19 cases was obtained from the COVID-19 Canada Open Data Working Group. We conducted time-lagged cross-correlation analyses and developed the long short-term memory model for forecasting daily COVID-19 cases. **Results:** Among symptom keywords, "cough," "runny nose," and "anosmia" were strong signals with high cross-correlation coefficients >0.8 ($r_{Cough} = 0.825, t-9$; $r_{Runny\ Nose} = 0.816, t-11$; $r_{Anosmia} = 0.812, t-3$), showing that searching for "cough," "runny nose," and "anosmia" on GT correlated with the incidence of COVID-19 and peaked 9, 11, and 3 days earlier than the incidence peak, respectively. For symptoms- and COVID-related Tweet counts, the cross-correlations of Tweet signals and daily cases were $r_{Tweet\ Symptoms} = 0.868, t-11$ and $r_{Tweet\ COVID} = 0.840, t-10$, respectively. The LSTM forecasting model achieved the best performance ($MSE = 124.78, R^2 = 0.88, adjusted\ R^2 = 0.87$) using GT signals with cross-correlation coefficients >0.75. Combining GT and Tweet signals did not improve the model performance. **Conclusion:** Internet search engine queries and social media data can be used as early warning signals for creating a real-time surveillance system for COVID-19 forecasting, but challenges remain in modelling.

Keywords. Emerging infectious diseases, COVID-19, digital surveillance systems, internet search engines, social media, Google Trends, Twitter, long short-term memory (LSTM)

1. Introduction

Emerging infectious diseases (EIDs), including COVID-19, Ebola virus, and SARS, pose significant threats to global public health and economies. As of March 1, 2023, the global

[1] Corresponding Author: Yang Yang, E-mail: y24yang@uwaterloo.ca.

confirmed cases of COVID-19 stand at approximately 675.4 million, with roughly 6.8 million deaths reported [1]. The economic costs of EIDs, such as COVID-19, are also staggering. According to the Canadian Institute for Health Information (CIHI), in 2021, due to the impact of the COVID-19 pandemic, Canada's healthcare costs were expected to reach an all-time high of $308 billion. This translates to roughly $580,000 per minute, with healthcare expenses accumulating to around $70,000 in the time it takes to read this sentence, or approximately $10,000 per second [2]. During global epidemics, the general public tends to utilize media, such as internet searches and social media, for health information more frequently than ever. This heightened consumption of online and social media data can facilitate real-time surveillance of emerging diseases and the prediction of epidemics [3-4]. Our team conducted a scoping review on COVID-19, which identified a gap in research and implementation of a real-time surveillance system for COVID-19 [5]. An efficient real-time surveillance system would serve as an early warning system, empowering public health authorities and hospitals to respond quickly to EID threats. Google Trends (GT) and Twitter can be used for digital disease surveillance in real-time, such as the HealthMap system [6], which uses machine learning algorithms to track outbreaks, and the Pandemic Response Platform [7] developed by the WHO and Microsoft to track disease spread and provide real-time updates to public health officials. This study aims to evaluate the association between internet search engine queries such as GT and social media data on Twitter about COVID-19 in Canada and investigate if information from these sources has predictive power as early warning signals to predict COVID-19 cases.

2. Methods

2.1. Data Collection and Preprocessing

We extracted GT and Twitter data from 1/1/2020 to 3/31/2020 in Canada as early warning signals, including symptom keywords (‘cough,’ ‘runny nose,’ ‘anosmia,’ ‘sore throat,’ ‘shortness of breath,’ ‘fever,’ ‘headache,’ ‘body ache,’ ‘dyspnea,’ and ‘fatigue’) from GT and COVID-19-related hashtags (‘pneumonia,’ ‘cough,’ ‘fever,’ ‘running nose,’ and ‘breath’) from Twitter, respectively. COVID-19 data was obtained from the COVID-19 Canada Open Data Working Group. All data were normalized to the same scale for analysis.

2.2. Data Analysis

We applied Fast Fourier Transform (FFT), Moving Average (MA), Savitzky–Golay (SG), and Lowess smoothing methods to remove the white noise obtained in the data. Denoising was done using adjacent averaging on both GT and Tweet signals. We performed time-lagged cross-correlation analyses using denoised signals to examine the relationship between each signal and daily COVID-19 cases. We then developed the long short-term memory (LSTM) model for forecasting COVID-19 cases with TensorFlow and Keras in Python. We used the past 20 days’ data for model training. The LSTM models were trained using the Adam optimizer, and the hidden units were 148 through experiments. We calculated the loss function, mean squared errors (MSE), and R^2 and adjusted R^2 to evaluate the model performance. We further fine-tuned the initial learning rate and learning rate drop hyperparameters, ranging between 0.0001 and 0.001.

3. Results

3.1. Time-Lagged Cross-Correlation Analysis

The time-lagged cross-correlation analysis between GT signals and daily COVID-19 cases showed a range of [1, 13] days of time lag. The maximum correlation coefficients varied greatly, ranging between -0.275 (Fatigue) and 0.825 (Cough). The cross-correlation analysis between Tweet signals and daily COVID-19 cases showed similar lags and coefficients. Table 1 summarizes the cross-correlation analysis results. Figure 1 shows the implementation of denoising the GT data using the MA method (left) and the denoised GT and Tweet signals (right).

Table 1. Time-lagged cross-correlation analyses on Google Trends and Twitter data.

Signals	Lag (in days from time t)	Max Correlation Coefficient
	Google Trends (GT)	
Cough	-9	0.825
Runny Nose	-11	0.816
Anosmia	-3	0.812
Sore Throat	-6	0.790
Shortness of Breath	-9	0.762
Fever	-10	0.752
Headache	-8	0.723
Body Ache	-5	0.612
Dyspnea	-13	0.501
Fatigue	-1	-0.275
	Tweet	
Symptom related	-11	0.868
COVID-19 related	-10	0.840

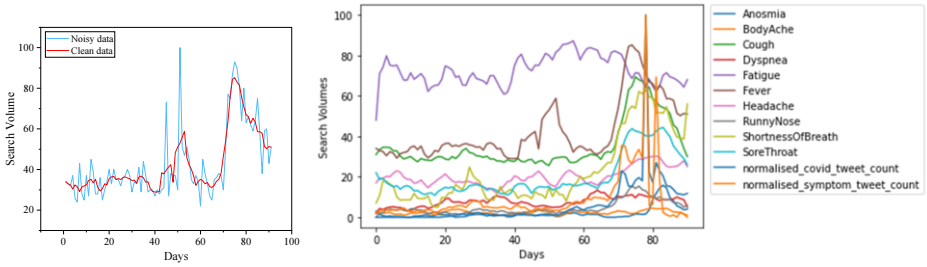


Figure 1. Signal processing. **Left**: Noisy and clean data for Google Trends for the keyword 'Fever.' **Right**: Denoised Google Trends and Tweet signals from 1/1/2020 to 3/31/2020 in Canada.

3.2. LSTM Forecasting Models

We experimented with different subsets of features using GT, Tweet, and the combination of GT and Tweet signals for LSTM forecasting models. Using GT signals with a correlation coefficient >0.75 achieved the best model performance ($MSE = 124.78, R^2 = 0.88, adjusted\ R^2 = 0.87$). Table 2 compares LSTM modelling results.

Table 2. Long short-term memory (LSTM) modelling results. **M1**: 'Cough,' 'Runny Nose,' 'Anosmia,' 'Sore Throat,' 'Shortness of Breath,' 'Fever,' 'Headache,' 'Body Ache.' **M2**: 'Cough,' 'Runny Nose,' 'Anosmia,' 'Sore Throat,' 'Shortness of Breath,' 'Fever,' 'Headache.' **M3**: 'Cough,' 'Runny Nose,' 'Anosmia,' 'Sore Throat,' 'Shortness of Breath,' 'Fever.' **M4**: 'Cough,' 'Runny Nose,' 'Anosmia.' **M5**: Tweet symptoms-related counts. **M6**: Tweet symptoms-related counts, Tweet COVID-related counts. **M7**: M3+M5. **M8**: M3+M6. **MSE**: mean squared error. **Adj R^2**: adjusted R^2.

Model	M1	M2	M3	M4	M5	M6	M7	M8
MSE	413.40	495.35	124.78	496.62	23500.5	26611.8	325.57	365.42
R^2	0.59	0.51	0.88	0.51	-22.89	-26.06	0.68	0.64
Adj R^2	0.57	0.48	0.87	0.48	-23.81	-27.10	0.66	0.62

Figure 2 illustrates the performance of the best model, M3.

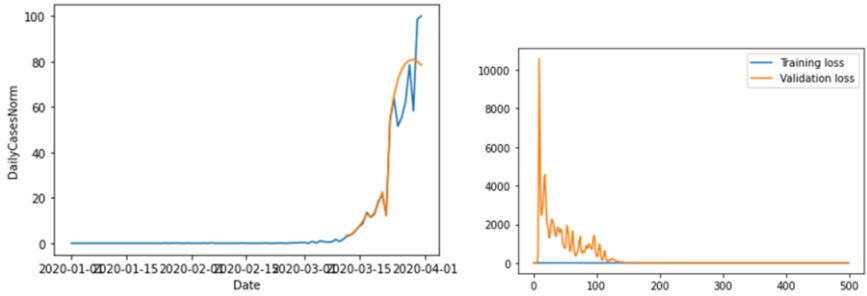


Figure 2. The long short-term memory (LSTM) model uses Google Trends signals with a maximum correlation coefficient >0.75. **Left**: COVID-19 confirmed cases in Canada (blue line: true values; orange line: predicted values). **Right**: training and validation loss (x-axis: epochs; y-axis: loss as measured by mean squared errors)

4. Discussion

This study uses denoised internet search engine queries (GT) and social media (Twitter) data to predict COVID-19 cases in Canada, demonstrating that these sources can be used for digital disease surveillance as an early warning system for EIDs. The choice of denoising method depends on the characteristics of the data and desired accuracy and reliability. FFT was found to be difficult to use on our data due to the lack of character frequencies in the white noise. MA, SG, and Lowess worked better than FFT and are suitable for data with a consistent trend, though they may need to be more effective for data with sudden changes in trend. GT data had more noise than Tweet data, possibly due to the normalized nature of GT data.

Using time-lagged cross-correlation analysis, this study found that COVID-19 symptom-related search terms strongly correlated with daily COVID-19 cases with a time lag of 1-13 days. GT and Twitter data can serve as early warning signals for real-time digital disease surveillance. Still, the LSTM model using GT data performed better than the model using Twitter data or a combination of both. Although tweets can be used as early warning signals for COVID-19 outbreaks, they are reactive and less effective than GT in predicting cases. According to a study, a correlation has been observed between such data and the incidence of COVID-19, with the data peaking 10-14 days before the incidence peak [8]. Other social media platforms like Facebook [9] and Reddit [10] can also provide valuable data for real-time disease surveillance for EIDs. Our study highlights the potential of internet search engines and social media data for real-time

disease surveillance. Still, challenges, such as limited data at the beginning of the COVID-19 pandemic, changing conditions, and unforeseen events, can impact accuracy. In addition, social media data may be subject to biases, such as the under-representation of certain groups, and may not always capture the full picture of an EID outbreak.

Strengths of this study include utilizing signal processing techniques and LSTM modelling to analyze internet search queries and social media data. However, GT data has a lower resolution and only provides relative search volume, and this study only used English Tweets, which may exclude valuable information. Additionally, there may be issues with the accuracy and reliability of Twitter's geolocation. Furthermore, retrospective time-series analyses are useful for monitoring disease outbreaks but can be prone to over-fitting, especially for complex data with many variables. Internet and social media data can provide valuable insights but should be considered alongside other surveillance sources to ensure a comprehensive understanding of disease trends. A multi-faceted strategy, such as incorporating multiple data sources and multimodal modelling with infectious disease models, is warranted for accurate and comprehensive EID surveillance. Lastly, identifying relevant symptom keywords of an EID requires a dynamic and adaptive approach that integrates various information sources and is continually refined as new data becomes available. Our future research will employ an ontology-based method to systematically identify and organize pertinent symptom keywords for an EID, even before they are commonly recognized or reported.

5. Conclusions

A real-time digital disease surveillance system that utilizes internet search engine queries and social media data can be an early warning system for forecasting EIDs like COVID-19. Such a system has the potential to assist in epidemiological control and monitor public perceptions of the disease, as well as forecast trends in outbreaks. However, challenges in modelling arise due to the noise of self-generated data and the identification of relevant symptom keywords of an EID before they are publicly known.

References

[1] Johns Hopkins University and Medicine. Coronavirus resource center: world map. https://coronavirus.jhu.edu/map.html. Date accessed: January 5, 2023
[2] COVID-19 expected to push Canada's health spending beyond $300 billion in 2021. https://www.cihi.ca/en/news/covid-19-expected-to-push-canadas-health-spending-beyond-300-billion-in-2021 Date accessed: December 8, 2022
[3] Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. Journal of medical Internet research. 2020 Apr 21;22(4):e19016.
[4] Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review. Journal of biomedical informatics. 2016 Aug 1;62:1-1.
[5] Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. The Lancet Digital Health. 2021 Mar 1;3(3):e175-94.
[6] About HealthMap https://healthmap.org/about/. Date accessed: December 8, 2022
[7] Pandemic Prevention Platform (P3) https://www.darpa.mil/program/pandemic-prevention-platform Date accessed: December 8, 2022
[8] Li C, et al. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance. 2020 Mar 12;25(10):2000199.
[9] Hossain L, Kam D, Kong F, Wigand RT, Bossomaier T. Social media in Ebola outbreak. Epidemiology & Infection. 2016 Jul;144(10):2136-43.
[10] Hu M, Conway M. Perspectives of the COVID-19 Pandemic on Reddit: Comparative Natural Language Processing Study of the United States, the United Kingdom, Canada, and Australia. JMIR infodemiology. 2022 Sep 27;2(2):e36941.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230291

Using Big Data to Uncover Association Between Sildenafil Use and Reduced Risk of Alzheimer's Disease

Xingyue HUO, MSPH[a,1] and Joseph FINKELSTEIN, MD, PhD[a]
[a]*Icahn School of Medicine at Mount Sinai, New York, NY, USA*
ORCiD ID: Xingyue Huo https://orcid.org/0000-0002-5135-0292

Abstract. Alzheimer's disease is a chronic neurodegenerative disease with multiple pathogenesis pathways. Sildenafil, one of the phosphodiesterase-5 inhibitors, was proven to have effective benefits in transgenic Alzheimer's disease mice. The purpose of the study was to investigate the relationship between sildenafil use and the risk of Alzheimer's disease based on the IBM® MarketScan® Database covering over 30 million employees and family members per year. Sildenafil and non-sildenafil-matched cohorts were generated using propensity-score matching with the greedy nearest-neighbor algorithm. The propensity score stratified univariate analysis and the Cox regression model showed that sildenafil use was significantly associated with a 60% risk reduction of developing Alzheimer's disease (HR=0.40; 95%CI:0.38-0.44; P<.0001) compared to the cohort of individuals who did not take sildenafil. Sex-stratified analyses revealed that sildenafil was related to a lower risk of Alzheimer's disease in subgroups of both males and females. Our findings demonstrated a significant association between sildenafil use and a lower risk of Alzheimer's disease.

Keywords. Public health, Alzheimer's disease, Sildenafil, Big data, Epidemiology informatics

1. Introduction

Alzheimer's disease (AD), a chronic neurodegenerative disease with multiple pathogenesis pathways, is the seventh leading cause of death in the United States. Recent studies estimated that AD could lead to dementia in more than 6 million Americans, most of whom are 65 or older. AD is a progressive brain disorder whose symptoms worsen over time and cannot be stopped or reversed. Despite significant advancements in our understanding of AD diagnosis and treatment, AD patients desperately need novel therapeutics to slow the progression of this life-threatening disease.

Sildenafil, a phosphodiesterase (PDE) inhibitor, is an oral medication approved by the US Food and Drug Administration for erectile dysfunction (ED) and pulmonary atrial hypertension (PAH). Sildenafil treats PAH by relaxing the blood vessels in the lungs to increase blood flow, while it also treats ED by increasing blood flow to the penis during sexual stimulation.

[1] Corresponding Author: Xingyue Huo, E-mail: hxy6879@hotmail.com.

It has been found that the temporal cortices of AD patients have significantly increased PDE5 protein levels. A number of studies proved that PDE inhibitors regulate signaling pathways by increasing cAMP and/or cGMP levels. The PDE5 inhibitors could potentially interrupt the NO/cGMP/PKG/CREB signaling pathway by increasing the levels of cGMP [1-2]. PDE5 inhibition has been proposed as a potential therapeutic approach for treating AD. Repurposing existing drugs could potentially result in more effective and accessible therapies for AD patients [3]. The purpose of the study was to investigate the relationship between sildenafil use and a lower risk of Alzheimer's disease using a large database of commercial healthcare claims.

2. Methods

2.1. Study Design

The study utilized the IBM©,R MarketScan©,R Medicare Supplemental Database from 2016 to 2019 containing information on the majority of commercially insured individuals in the U.S. Sildenafil prescriptions were identified using the National Drug Code. AD, as well as other medical diagnoses, were derived from individuals' medical records and determined by the international classification of diseases 10th revision (ICD-10) code. All study participants were required to adhere to the following requirements: 1) they should have continuous insurance coverage from 2016 to 2019; 2) they were 65 years or older at the start of the study; 3) they have never been diagnosed with mild cognitive impairment (MCI), encephalopathy, or dementia of any type; 4) the AD diagnosis date should be after the date the drug under investigation was prescribed.

The participants in the study who received a prescription for Sildenafil were referred to as the sildenafil exposure cohort, while the participants who did not receive a prescription for any medication were referred to as the control cohort. The study's outcome measured the interval between the start of sildenafil use and the diagnosis of AD. Subjects without AD diagnosis were censored from the cohorts at the end of the study period.

The covariates included the Charlson Comorbidity Index (CCI), age, sex, and the U.S. region. The CCI was categorized based on comorbidities in the patient's medical records. Each comorbidity category has a corresponding weight, and the total of weights yields a patient's overall comorbidity score. A higher score indicates a greater likelihood of higher morbidity and mortality. A score of 0 means that no comorbidities were found. As a result, we divided CCI into four groups: 0, 1, 2-3, and greater than or equal to 4. Age was categorized into groups of 65 to 74 and 75 or older.

2.2. Statistical analysis

The propensity-score matching (PSM) method was used to control the confounding factors. First, a propensity score of sildenafil exposure cohorts was calculated for each patient using multivariable logistic regression, adjusting for sex, age, geographic region, and CCI. Then, we performed greedy PSM and generated the matched groups at a ratio of 1:1 using the nearest neighbor within the caliper matching strategy. We set a maximum distance within matches with a caliper of 0.2 standard deviations to improve

the matching performance. Every patient from the sildenafil cohort was matched with a patient from the non-sildenafil cohort who shared the same gender, age, region, and CCI.

Univariate analysis of the demographic variables was used to identify potential significant variables that could be used in model runs. All variables that had a chi-square value or a t statistic associated to a p value of 0.15 or less were considered as potentially significant and retained for the model analysis. Additionally, PS-stratified Cox proportional hazards models were utilized to perform the hazard ratios (HRs) of developing AD between sildenafil cohorts and non-sildenafil cohort.

All analysis were performed using procedures in SAS version 9.4 (SAS Institute, Cary, NC). Two-sided p-value of < 0.05 was considered as statistically significant.

3. Results

All medical record data was collected from the year 2016 to 2019 IBM® MarketScan® Medicare claims Databases. Overall, 17,125 (32.4%) individuals had at least one sildenafil prescription, while 35,652 (67.6%) had never received one. After performing 1:1 PMS, there were 13,587 individuals in each of the sildenafil cohort and non-sildenafil cohort. Out of the 27,174 participants in the PSM-matched cohorts (Table 1), 5,674 (10.7%) were diagnosed with AD. Sildenafil users had a lower risk of developing AD (5.7%) than non-sildenafil users, who had an AD incidence of 9.5%.

Table 1. Description of Characteristics for Univariate Analysis

		No AD diagnosis	AD diagnosis	
Characteristics		N=47,103(89.3%)	N=5,674 (10.7%)	**P-value**
Sildenafil	Sildenafil	12,812 (94.3%)	775 (5.7%)	<.0001
	Non-sildenafil	12,291 (90.5%)	1,296 (9.5%)	
Charlson Comorbidity Index Score	0	5,004 (96.5%)	181 (3.5%)	<.0001
	1	5,451 (94.6%)	309 (5.4%)	
	2-3	8,885 (91.8%)	797 (8.2%)	
	>=4	5,763 (88.0%)	784 (12.0%)	
Sex	Male	24,795 (92.5%)	2,015 (7.5%)	<.0001
	Female	308 (84.6%)	56 (15.4%)	
Region	NE	12,142 (91.7%)	1,097 (8.3%)	<.0001
	NC	3,569 (92.1%)	306 (7.9%)	
	South	7,583 (93.3%)	546 (6.7%)	
	West	1,809 (93.7%)	122 (6.3%)	
Age	65-74 yrs	18,110 (95.0%)	963 (5.1%)	<.0001
	>=75 yrs	6,993 (86.3%)	1,108 (13.7%)	

We found that sildenafil was significantly associated with a 60% lower risk of developing AD (HR =0.40; 95% CI: 0.38-0.44; P<.0001) (Table 2) in comparison to a cohort of individuals who did not take sildenafil. Particularly, sex-stratified analyses revealed that sildenafil was related to a lower risk of AD in subgroups of both men and

women. According to a crude Cox regression model, sildenafil is found to be significantly associated with a 62% lower risk of AD in men (HR=0.38; 95% CI: 0.35-0.42; P<.0001), a 47% lower risk of AD in women (HR=0.53; 95% CI: 0.31-0.90; P=0.019), and a 62% lower risk of AD across all genders (HR=0.38; 95% CI: 0.35-0.42; P<.0001). The outcome of the Cox regression model with the adjusted CCI score and age is also reliable.

Table 2. Hazard ratios (HR) and 95% CI in unadjusted model and model after adjusting for age, region, CCI score.

	Sex=Male				Sex=Female				Sex= All			
Parameter	**HR**	**95% CI**		**P-Value**	**HR**	**95% CI**		**P-Value**	**HR**	**95% CI**		**P-Value**
Unadjusted model												
Sildenafil vs. Non-drug exposure	0.38	0.35	0.42	<.0001	0.53	0.31	0.9	0.02	0.38	0.35	0.42	<.0001
Adjusted model												
Sildenafil vs. Non-drug exposure	0.40	0.36	0.44	<.0001	0.47	0.27	0.81	0.01	0.40	0.36	0.44	<.0001
65-74 yrs vs. >=75 yrs	0.38	0.35	0.41	<.0001	0.10	0.04	0.23	<.0001	0.38	0.35	0.41	<.0001
1 vs. 0 (No comorbidity)	1.42	1.18	1.70	0.0002	0.47	0.03	7.48	0.59	1.42	1.18	1.70	0.0002
2-3 vs. 0 (No comorbidity)	2.10	1.79	2.47	<.0001	2.96	0.39	22.33	0.29	2.10	1.79	2.47	<.0001
>=4 vs. 0 (No comorbidity)	3.08	2.62	3.63	<.0001	1.83	0.25	13.66	0.56	3.08	2.62	3.63	<.0001

4. Discussion

We analyzed the association between the sildenafil use and the risk of AD in older adults aged 65 years or older and then found that sildenafil could reduce the risk of AD by 60%. Our study also suggests that sildenafil could be a promising potential candidate drug in the treatment of AD. More randomized clinical trials of sildenafil in patients with AD are warranted in the future.

Several studies have found that PDE5 inhibitors can modestly increase cerebral blood flow at rest in patients with cognitive impairment but did not adequately address the effects of sildenafil in AD [4-5]. In addition, sildenafil is an effective medication for treating ED. As per studies, the risk factors for older men who need sildenafil to treat ED are cardiovascular disease, hypertension, diabetes mellitus, tobacco use, hyperlipidemia, metabolic syndrome, and depression [6]. The recent meta-analysis of the potential role of PDE-5 inhibitors in the treatment of dementia corroborated our results [9].

Therefore, several analyses need to be conducted regarding sildenafil and AD in different study cohorts [7]. Cox regression analyses were successfully used to estimate the hazard rate (HR) for use of ED medications, controlled for multiple time-varying covariates in previous studies [7-8]. In addition, we plan to conduct further analyses to

see if other FDA-approved medications in the PED 5 inhibitor class, such as Tadalafil, may have comparable effects in lowering the risk of Alzheimer's disease.

5. Conclusion

This is the first study that used MarketScan, a nationwide database containing over 30 million insured members annually, to assess the association between sildenafil usage and Alzheimer's disease. Our approach's novelty is based on applying the unique big data resource to test the hypothesis that sildenafil may have protective properties against Alzheimer's disease. Previous retrospective studies used much smaller samples and produced contradicting results [3, 7-8]. Inclusion of the Charlson Comorbidity Index helped with accounting for multiple comorbidity burden in the target population. After adjusting for the relevant confounders, the Cox regression model showed that sildenafil use was significantly associated with a 60% risk reduction of developing Alzheimer's disease.

References

[1] Sanders O. Sildenafil for the treatment of Alzheimer's disease: A systematic review. Journal of Alzheimer's Disease Reports. 2020 Jan 1;4(1):91-106.

[2] Kang BW, Kim F, Cho JY, Kim S, Rhee J, Choung JJ. Phosphodiesterase 5 inhibitor mirodenafil ameliorates Alzheimer-like pathology and symptoms by multimodal actions. Alzheimer's Research & Therapy. 2022 Dec;14(1):1-7.

[3] Fang J, Zhang P, Zhou Y, Chiang CW, Tan J, Hou Y, Stauffer S, Li L, Pieper AA, Cummings J, Cheng F. Endophenotype-based in silico network medicine discovery combined with insurance record data mining identifies sildenafil as a candidate drug for Alzheimer's disease. Nature Aging. 2021 Dec;1(12):1175-88.

[4] Sheng M, Lu H, Liu P, Li Y, Ravi H, Peng SL, Diaz-Arrastia R, Devous MD, Womack KB. Sildenafil improves vascular and metabolic function in patients with Alzheimer's disease. Journal of Alzheimer's Disease. 2017 Jan 1;60(4):1351-64.

[5] Samudra N, Motes M, Lu H, Sheng M, Diaz-Arrastia R, Devous M, Hart J, Womack KB. A pilot study of changes in medial temporal lobe fractional amplitude of low frequency fluctuations after sildenafil administration in patients with Alzheimer's disease. Journal of Alzheimer's Disease. 2019 Jan 1;70(1):163-70.

[6] DeLay KJ, Haney N, Hellstrom WJ. Modifying risk factors in the management of erectile dysfunction: a review. The world journal of men's health. 2016 Aug;34(2):89-100.

[7] Desai RJ, Mahesri M, Lee SB, Varma VR, Loeffler T, Schilcher I, Gerhard T, Segal JB, Ritchey ME, Horton DB, Kim SC. No association between initiation of phosphodiesterase-5 inhibitors and risk of incident Alzheimer's disease and related dementia: results from the Drug Repurposing for Effective Alzheimer's Medicines study. Brain Communications. 2022;4(5):fcac247.

[8] Friedman S, Magnussen B, O'Toole A, Fedder J, Larsen MD, Nørgård BM. Increased use of medications for erectile dysfunction in men with ulcerative colitis and Crohn's disease compared to men without inflammatory bowel disease: a nationwide cohort study. Official journal of the American College of Gastroenterology| ACG. 2018 Sep 1;113(9):1355.

[9] Lee DH, Lee JY, Hong DY, Lee EC, Park SW, Jo YN, Park YJ, Cho JY, Cho YJ, Chae SH, Lee MR. ROCK and PDE-5 inhibitors for the treatment of dementia: literature review and meta-analysis. Biomedicines. 2022 Jun 8;10(6):1348.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230292

Conducting an Epidemiologic Study and Making It FAIR: Reusable Tools and Procedures from a Population-Based Cohort Study

Carsten Oliver SCHMIDT[a,1], Stephan STRUCKMANN[a], Maik SCHOLZ[a], Janka SCHÖSSOW[a], Dörte RADKE[a], Adrian RICHTER[a], Achim REINEKE[b], Elisa KASBOHM[a], Joany Mariño CORONADO[a], Birgit SCHAUER[a], Kristin HENSELIN[a], Susanne WESTPHAL[a], Darko BALKE[a], Torsten LEDDIG[a], Henry VÖLZKE[a] and Jörg HENKE[a]

[a] *University Medicine Greifswald, Institute for Community Medicine. Greifswald, Germany*

[b] *Leibniz Institute for Prevention Research and Epidemiology - BIPS. Bremen, Germany*

ORCiD ID: C Schmidt (0000-0001-5266-9396), M Scholz (0009-0000-5385-864X), S Struckmann (0000-0002-8565-7962), J Schössow (0009-0003-4598-3397), A Richter (0000-0002-3372-2021), D Radke (0009-0006-4926-8018), A Reineke (0000-0002-0092-4110), E Kasbohm (0000-0001-5261-538X), J Mariño (0000-0002-4657-3758), B Schauer (0000-0001-9847-130X), K Henselin (0009-0004-6290-8468), S Westphal (0009-0008-7580-8465), D Balke (0009-0003-1550-0470), T Leddig (0000-0001-8883-5480), H Völzke (0000-0001-7003-399X), J Henke (0009-0000-3583-4310)

Abstract. Conducting large-scale epidemiologic studies requires powerful software for electronic data capture, data management, data quality assessments, and participant management. There is also an increasing need to make studies and the data collected findable, accessible, interoperable, and reusable (FAIR). However, reusable software tools from major studies, underlying such needs, are not necessarily known to other researchers. Therefore, this work gives an overview on the main tools used to conduct the internationally highly networked population-based project Study of Health in Pomerania (SHIP), as well as approaches taken to improve its FAIRness. Deep phenotyping, formalizing processes from data capture to data transfer, with a strong emphasis on cooperation and data exchange have laid the foundation for a broad scientific impact with more than 1500 published papers to date.

Keywords. FAIR, cohort study, web applications, quality management, data quality

[1] Corresponding Author: Carsten Oliver Schmidt, University Medicine Greifswald, Institute for Community Medicine, SHIP-Clinical Epidemiological Research – Functional Division Quality in the Health Sciences, Walther Rathenau Str. 48, 17475 Greifswald, Germany; Email: Carsten.schmidt@uni-greifswald.de.

1. Introduction

Successfully running an epidemiologic study requires a wide range of IT-tools to cover aspects such as electronic data capture, data management, quality management, and participant management. There is a growing need to make studies FAIR to facilitate networked research and increase reproducibility [1]. Meeting all demands is a complex challenge in large-scale epidemiologic cohort studies [2]. Potentially powerful solutions are in place to meet demands from such studies. These solutions often remain unknown to other scientists and groups, despite their potential for reuse. This work therefore provides an overview of software tools and processes that have been put in place to conduct an internationally highly networked cohort-study and make it FAIR.

2. Methods

2.1. The Study of Health in Pomerania (SHIP)

The SHIP project studies the prevalence and incidence of risk factors, subclinical disorders, clinical diseases, and their inter-relations [3]. It consists of three cohorts (SHIP-START: N=4308; SHIP-TREND: N=4420; SHIP-NEXT: baseline ongoing, projected N=4000), the first of which started in 1997. Data collections are ongoing in all three cohorts. Major follow-up examinations are carried out in approximately five-year intervals. SHIP has implemented one of the widest scope of examinations worldwide in population-based research, including amongst others interviews, questionnaires, biomaterials (blood, urine, faeces, saliva), imaging (e.g., ultrasound; full-body magnetic resonance imaging (MRI)), cardiovascular, dental, and dermatological examination, polysomnography, body scanning, as well as examinations of animals. Not counting OMICS data and externally linked data sources, the SHIP database includes far more than 50,000 data elements. Attached to the main examination waves are hundreds of side projects (e.g., MRI readings) with independent but centrally managed data collections.

2.2. IT-background

SHIP web-applications have been developed using, amongst others, the technology of Java Server Faces. They are partly in a refactoring process to migrate to the Spring Boot architecture. All of them are provided as Tomcat (https://tomcat.apache.org) applications. GitLab is used as a development infrastructure that includes a central source code repository for distributed version control, a wiki for documentation, and a ticket system. From the beginning, internationalization was implemented to enable the handling of different languages. As part of continuous integration, some SHIP applications are build automatically for SHIP's own repositories. Auto deployment can be done easily this way. Keycloak is currently being established to allow for a single sign-in. All SHIP data are stored in a central PostgreSQL database.

3. Results

3.1. Tools for conducting the study and quality management

Table 1 provides a brief description of the major tools used to run SHIP data collections, accompanying processes such as data and quality management, as well as participant management. Figure 1 sketches their interrelation. Many tools have already been reused in other major studies. Examples are SHIPPIE and SHIPdesigner in the polish Bialystok PLUS study, WebMODYS and Square² in the German National Cohort (NAKO Gesundheitsstudie), the largest German epidemiologic cohort study to date, dataquieR in Eurocrine. The following paragraphs describe approaches to improve the FAIRness of SHIP.

3.2. Improving Findability

To make content better findable rather than relying on study specific web pages, several collaborations have been established, e.g. with Maelstrom Research [4] (https://www.maelstrom-research.org/), the Portal of Medical Data Models [5] (https://medical-data-models.org/), NFDI4Health (https://www.nfdi4 health.de/), and euCanSHare (https://mica.eucanshare.bsc.es/) to facilitate the search for SHIP data. A key feature of these collaborations has been the semantic annotation of data elements by cooperation partners to describe the content of study variables, using either the Maelstrom taxonomy (Maelstrom, euCanSHare, NFDI4Health) or Unified Medical Language System (UMLS, MDM). This greatly enhances comparison options with other studies, such as UK-Biobank or Rotterdam Study, amongst others.

3.3. Improving Accessibility

For legal and data protection issues, access to SHIP data cannot directly be made available publicly or from a central portal. However, a dedicated public web-site (https://www.fvcm.med.uni-greifswald.de/) and formal Use & Access process enables sustainable access to SHIP data [3]. More than 200 applications have been processed annually in the previous years. The transfer-site is open internationally for scientists to register for data requests.

3.4. Improving Interoperability

The SHIP PostgreSQL database has a proprietary data model but the need for exchange with other Central Metadata Repositories has led to the creation of export routines to formats such as OPAL, CDISC, as well as common formats used in statistical software such as R, SAS, Stata, and SPSS. Availability of SHIP forms via MDM [5] also enables their direct export and use in applications like REDCap.

3.5. Improving Reusability

Since study data and metadata are stored in a central database, they can be easily managed in terms of data protection, availability and long-term archiving. The processes are supported by a fully automated pipeline that covers all steps of data management

from data capture, through data curation to data extraction for scientists with accepted data applications [6]. Reusability also requires high quality data and metadata. Therefore, extensive data quality assessment pipelines have been implemented to find and fix issues.

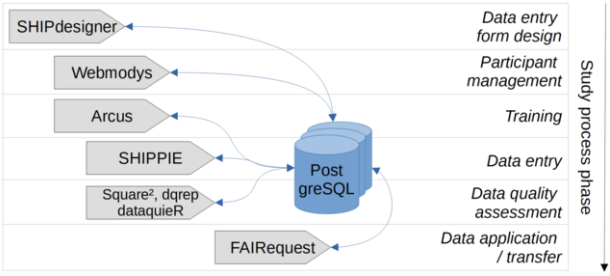


Figure 1. Position of SHIP tools in different phases of the study life cycle

Table 1. Overview of SHIP tools for conducting studies and quality management

SHIPPIE is used for electronic data capture, using Java Server Faces, Jooq, and a PostgreSQL database. SHIPPIE generates web forms in a fully dynamic manner based on specifications defined using SHIPdesigner. SHIPPIE has been designed with demands to handle multiple examinations and visits per person. A more extensive set of metadata information than usual is in place to enable automated data quality assessments. Not only simple limit violations, but also conditional violations can be assessed by entering Boolean expressions in the disjunctive normal form. Measurements from previous observations can be displayed as needed to prevent input errors. The web application offers two modes: the recording mode and the test mode. In the latter, recorded data are stored in a test database table. This way, no second test installation is needed for testing purposes. A powerful rights-and-role system allows to specify users' access rights according to their role in the study.
SHIPdesigner supports SHIPPIE users in creating and editing web forms. It uses primefaces (https://www.primefaces.org/) to render the GUI rather than standard jsf (plain html) components. Therefore, neither new database tables nor updates of the SHIPPIE application are needed, if forms are added. This increases the implementation speed of changes in the web forms and decouples the operational use of the applications from the development process.
WebMODYS is a web-application to support, control, and document participant management in population-based studies. It has been developed as a Java web application in a cooperation with BIPS, Bremen, Germany. A prominent feature lies in its free configurability for study-specific recruitment processes, which is achieved by storing the recruitment processes as state diagrams in the database. WebMODYS provides functions for daily participant management tasks and their documentation like editing and updating basic participant data, generating letters, scheduling appointments, and logging all contacts between participants and recruitment personnel as well as other participant-related recruitment tasks. It enables a full protocol of any participant contact and provides data to calculate response proportions or to analyze non-response.
ARCUS is a web application to train and certify readers of imaging data and is used for example in the context of ultrasound examinations. It was programmed using Java Server Faces, and PostgreSQL, and can read DICOM images in 16-bit color depth and allows measurements directly on those images. Arcus issues training-certificates for readers who successfully completed their training.
dataquieR, **dqrep** and **Square²** were designed to enable an automated data quality monitoring. dataquieR is available as an R package [7], which can be downloaded from CRAN. dqrep is a Stata package with functionalities for multi-report generation. Square² is a JAVA web application [8]. Square² offers a graphical user interface (GUI) to target all steps in the data quality assessment workflow: implementing the study structure, managing needed metadata and finally creating quality reports.
FAIRequest refers to a web application in Spring Boot that has replaced an older php data-application tool to search and apply for SHIP data using standardized data application access forms. Variable browsing is possible through direct links to the SHIP data dictionary.

There are further tools in place such as a daily run fully automated modularized SAS data-cleaning pipeline. It's output is used by an Access application to collect feedback on detected issues [6]. The **JoinUs4Health** platform (https://platform.joinus4health.eu) is a Wordpress application that allows anybody from the age of 16 years to submit own suggestions for SHIP topics, or vote on contributions of others.

4. Discussion

This paper provides an overview of key tools to conduct SHIP and improve its FAIRness. Their use has helped make SHIP one of the most interconnected epidemiologic projects in the world, with hundreds of cooperation partners worldwide and more than 1500 peer-reviewed articles. Yet, there are also structural limitations to FAIRness, for example related to the public storage of highly sensitive personal health data in compliance with the EU-General Data Protection Regulation.
The tools mentioned here are publicly available and can be downloaded directly or used free of charge for academic purposes through collaborations. Some tools, such as SHIPPIE, have been tailored to the specific needs of complex cohort studies. Others are more general in nature, like the data quality and participant management software. Future challenges will be their gradual update, and a stronger integration with standards and developments such as FHIR or OMOP. We also aim to increase FAIRness of SHIP towards non-scientists such as participants, the general population, and stakeholders to promote citizen science through means such as the JoinUs4Health platform.

Declarations

Ethical vote: not applicable

Conflict of Interest: The authors declare that there is no conflict of interest.

Acknowledgement: Content of this work was funded by the Ministry for Education, Science and Culture of the State of Mecklenburg-Vorpommern, European Union's funds (No UG 11 035A , 825903, No. 101006518), and by the German Research Foundation (DFG, SCHM 2744/3-1; SCHM 2744/3-4; NFDI 13/1).

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018.
[2] Wichmann HE, Kuhn KA, Waldenberger M, et al. Comprehensive catalog of European biobanks. Nat Biotechnol 2011;29:795-7.
[3] Volzke H, Schossow J, Schmidt CO, et al. Cohort Profile Update: The Study of Health in Pomerania (SHIP). Int J Epidemiol 2022.
[4] Bergeron J, Doiron D, Marcon Y, Ferretti V, Fortier I. Fostering population-based cohort data discovery:. PLoS ONE 2018;13:e0200926.
[5] Hegselmann S, Gessner S, Neuhaus P, Henke J, Schmidt CO, Dugas M. Automatic conversion of metadata from the Study of Health in Pomerania to ODM. Stud Health Technol Inform 2017;236:88-96.
[6] Werner A, Maiwald S, Henselin K, et al. [Modular automated data cleaning in a large population-based cohort]. In: Chenot JF, Minkenberg R, eds. Proceedings of the 20 conference of SAS®-Users in research and development: Shaker; 2016:279-84.
[7] Richter A, Schmidt CO, Krüger M, Struckmann S. dataquieR: assessment of data quality in epidemiological research. JOSS 2021;6:3039.
[8] Schmidt CO, Krabbe C, Schössow J, Albers M, Radke D, Henke J. Square² - A web application for data monitoring in epidemiological and clinical studies. Stud Health Technol Inform 2017;235:549-53.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230293

Data Collection and Analysis Methods for Smart Nudging to Promote Physical Activity: Protocol for a Mixed Methods Study

Seshathiri DHANASEKARAN[a,1], Anders ANDERSEN[a], Randi KARLSEN[a], Anne HÅKANSSON[a] and André HENRIKSEN[a]
a Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway

Abstract. New digital technologies like activity trackers, nudge concepts, and approaches can inspire and improve personal health. There is increasing interest in employing such devices to monitor people's health and well-being. These devices can continually gather and examine health-related information from people and groups in their familiar surroundings. Context-aware nudges can assist people in self-managing and enhancing their health. In this protocol paper, we describe how we plan to investigate what motivates people to engage in physical activity (PA), what influences them to accept nudges, and how participant motivation for PA may be impacted by technology use.

Keywords. Personalized Nudge, Physical Activity, Activity Trackers, Data Analysis, Smart Watch

1. Introduction

A nudge is "any aspect of the choice architecture that alters people's behavior predictably without forbidding any options or significantly changing their economic incentives," according to Thaler and Sunstein's definition from 2008 [1]. Nudges were initially inves- tigated in offline decision-making, primarily focusing on financial or personal health choices. When the idea of nudging was applied to digital user interfaces in 2016, the term "digital nudging" was first introduced. Weinmass et al. [2] are credited with coin- ing the phrase and gave it the definition "subtle form of using design, information, and interaction elements to guide user behavior in digital environments, without restricting the individual's freedom of choice". Additionally, smart nudging was defined by Karlsen and Andersen [3] in 2019 as "digital nudging, where the guidance of user behavior is tailored to be relevant to the current situation of each individual user". A user profile with a large scope of connected data is necessary for smart nudging. Before a personalised nudge is created, the data is examined. In comparison to a non-personalized nudge, user acceptance of a customized nudge is

[1] Corresponding Author: Seshathiri Dhanasekaran, Department of Computer Science, UiT The Arctic University of Norway, Tromsø 9037, Norway. Email: seshathiri.dhanasekaran@uit.no.

more likely to succeed (i.e., the user accepts the nudge and follows the recommendation) [3]. Smartwatches, activity trackers, and smartphones can offer long-term, continuous behavioural data collecting. This comprises health-related data types like physical activity (PA), pulse, body temperature, stress levels, sleep, and contextual data (e.g., loca- tion by global positioning systems [GPS]) from which other data types can be inferred (e.g., weather). Thus, these devices offer a chance to create a continuous feedback loop to deliver timely cues, enabling people to better manage their own health and make wise decisions [4]. These devices may be able to provide insights into various demographic groups, with access to near real-time data collection and evaluation at the population level from a public health perspective, by gathering this type of data from a large portion of the population [5]. Understanding a person's context is challenging, and accurate smart device data are necessary to gain insightful knowledge [6]. Natural human conduct is influenced by several events occurring at once. People might be running, for instance, on running tracks, in the open air, and indoors [7]. In this circumstance, the geographic location information can help in determining the user's context. In this paper we present the protocol for an upcoming study, where the aim is to evaluate what motivates people to engage in physical activity and what influences them to accept nudges.

2. Methods

2.1. Study Context

This project is a part of High North Population studies (BiN). People living in Northern Norway are the primary population for this study. This study will help the high north residents to get relevant nudging to lead a healthy lifestyle. which aims to develop and test tailored nudges for people to increase physical activity to lead a healthy lifestyle. We have previously conducted a quantitative study on PA participation for such individuals [8,9]. The results from this study will be used in future smart nudging projects.

2.2. Participation Selection and Recruitment

Participants will be recruited by online invitation, posters, and by using convenience sampling. A minimum of 10 participants will be recruited for this study. All participants will be provided with an Apple watch series 8 and instructions will be provided for sharing their data to the researcher.

2.3. Inclusion and Exclusion Criteria

The inclusion criteria are: 1) resident of Northern Norway, 2) owning an Apple iPhone, 3) willing to wear an Apple watch for four weeks, 4) willing to share collected PA data, 5) willing to participate in one face-to-face interview, 6) age between 18 and 60 years, and 7) can provide written informed consent for participation prior to enrolment. The exclusion criteria are: 1) having an active gym subscription, 2) having a high level of PA, and 3) being unable to provide written informed consent.

2.4. Ethics Approval

The goal and methods of the study will be explained to each participant. Participants will be asked for written informed consent, as far as possible if the individual has the capacity to provide consent. The study description with a data management plan submitted for approval by the Regional Committees for Medical and Health Research Ethics in Norway and the data protection officer at UiT the Arctic University of Norway.

2.5. Data Collection

To access the data, we require participants to export the Apple Watch acquired data using the Apple Health application in their iPhone. Each participant's individual Extensible Markup Language (XML) file contains daily information on PA, workout specifics, and device information. The gathered information has combined a person's step count from their watch and mobile phone. We will not include steps taken from the mobile phone and only looks at step data from the watch. Using Python [10] with NumPy [11], we will extract primary variables from the original data and store in comma-separated value (CSV) files.
Daily variables for Active energy, Exercise minutes, Steps, and Distance will be kept in CSV files. Active energy is defined as the energy used by PA, expressed as Kcal. Light-, moderate-, and intensive PA minutes are combined to create exercise minutes. Distance is the sum of the running and walking distances, expressed in kilometres. The total number of daily steps recorded by the smartwatch is known as steps. Heart rate, heart rate variability (HRV), oxygen saturation (SpO2), peak oxygen uptake (VO2Max), stand minutes, and sleep information are additional variables [9]. We will collect this smart watch activity data from participants for a four week time period.

2.6. Qualitative Interview

After the 4-week data collection period, the recruited participants will be requested to take part in a 1-hour qualitative interview. The participant's home or other settings that are convenient for them, will be the site of a semi-structured interview [12]. We will prepare an interview guide that will be used during the interviews.
There will be two sections in the interview guide. Personal preferences and participant experiences with all areas of the study will be the main topics of Section 1. The influencing factors to improve PA levels will be the main topic of Section 2, along with practical challenges to nudging in everyday life and how the environment surrounding a person affects motivation for PA. All interviews will be recorded, verbatim transcribed, and anonymised.

3. Data Analysis

3.1. Quantitative Data

According to the nature and distribution of the data, appropriate quantitative statistical analysis will be carried out using Python (Python Software Foundation). The

descriptive statistics will be displayed as frequencies of categorical data, means with SDs or 95% CIs, and median (IQR) values. Estimates of effects utilising participant out- come measures will be investigated and published as estimates with 95% CIs without P values in accordance with the Consolidated Standards of Reporting Trials (CONSORT) 2010 extension [13,14]. We will also analyse the distributional characteristics of the variables. Additionally, we will do mixed analyses, where information from the qualitative content can be used to enhance quantitative data. The interviews can potentially help determine whether the outlying factors cause changes in activity data or if other causes are at play.

3.2. Qualitative Data

The interviews' transcripts will be examined using the tenets of systematic text condensation [15]. This analysis is conducted in two steps. Data on the viability of processes, expectations, and application use for PA, experiences with nudges, and motivation for acceptance of nudges to promote PA are provided through the qualitative approach. Clar- ity, explanation of minor themes, and a variety of cases will also be added to improve the conclusions' credibility and applicability. To ensure the greatest possible level of re- search quality in this pilot project, the consolidated standards for reporting qualitative research will be used [16].

4. Discussion

4.1 Expected Results and Findings

Enrolment will begin in January 2023 and when 10 participants have been included. Individual participant recruitment and intervention delivery will take place. This protocol's key contribution is a thorough explanation of a study that may reveal how people want themself to get motivated and whether they will accept being nudged to promote PA. This study will also investigate ways to persuade such people to adopt tailored nudging for a healthier lifestyle. This information can direct the creation of technology-based PA programmes in the future and enhance intervention studies that seek to raise PA levels. We anticipate that this study trial will reveal any potential issues with using technological tools and individual preferences for interventions meant to promote PA for those with low levels of exercise. This study will demonstrate how the application of technological use of data, like smart nudging, can be examined in personalized-context.

4.2. Limitation and Implications

There will be numerous restrictions on this investigation. The limited sample size will have an impact on how representative the quantitative results are. If data saturation is attained, a sample of 10 participants for the qualitative data is anticipated to be of high quality. The findings of the study could have significant effects on people. The research will be presented at conferences and released in reputable, peer-reviewed international journals.

5. Conclusion and Future Work

This study will assess the PA-influencing elements and acceptability of tailored nudges. The result from the study can be used to address problems with feasibility, enhance study methods, and calculate the effectiveness of the study's metrics. In addition, this paper investigates how technology with personalization can affect PA motivation in order to better inform and direct future technology-assisted PA interventions. To ensure improved health and a higher quality of life, it is crucial to investigate novel approaches to improve PA for individuals. We will use the data and outcome of this study to develop the machine learning-based future activity suggestion model for the same participants. We will use those outcomes to design and provide smart nudges for a larger population.

References

[1] Thaler RH, Sunstein CR. Nudge: Improving decisions about health, wealth, and happiness. Penguin; 2009 Feb 24.
[2] Weinmann M, Schneider C, Brocke JV. Digital nudging. Business & Information Systems Engineering. 2016 Dec;58(6):433-6.
[3] Karlsen R, Andersen A. Recommendations with a nudge. Technologies. 2019 Jun 13;7(2):45.
[4] Shaw RJ, Bonnet JP, Modarai F, George A, Shahsahebi M. Mobile health technology for personalized primary care medicine. The American journal of medicine. 2015 Jun 1;128(6):555-7.
[5] Henriksen A, Mikalsen MH, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. Journal of medical Internet research. 2018 Mar 22;20(3):e9157.
[6] Cherubini M, et al. The unexpected downside of paying or send- ing messages to people to make them walk: Comparing tangible rewards and motivational messages to improve physical activity. ACM Transactions on Computer-Human Interaction (TOCHI). 2020 Mar 20;27(2):1-44.
[7] Haga SB, et al. Promoting Wellness Through Mobile Health Technology in a College Student Population: Protocol Development and Pilot Study. JMIR Research Protocols. 2020 Apr 3;9(4):e16474.
[8] Dhanasekaran S, Andersen A, Karlsen R, Ha˚kansson A. Data Analysis Techniques for Smart Nudging. International Conference on AI-ML Systems 2021 Oct 22 (pp. 155-156) .
[9] Dhanasekaran S, Andersen A, Karlsen R, Ha˚kansson A, Henriksen A. Data collection and smart nudging to promote physical activity and a healthy lifestyle using wearable devices. InScandinavian Conference on Health Informatics 2022 Aug 22 (pp. 154-158).
[10] Python Software Foundation (2022), 'Python', (3.10.9 edn.: Guido van Rossum)
[11] The NumPy development team (2022), 'NumPy',(1.23.5 edn.: Sebastian Berg, Ralf Gommers, CharlesHarris, Stephan Hoyer, Melissa Weber Mendonc¸a,Inessa Pawson, Matti Picus, Ste´fan van der Walt, Eric Wieser)
[12] Michalsen H, et al. Family members and health care workers' perspectives on motivational factors of participation in physical activity for people with in- tellectual disability: A qualitative study. Journal of Intellectual Disability Research. 2020 Apr;64(4):259- 70.
[13] Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, Lancaster GA. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. bmj. 2016;355.
[14] Horne E, Lancaster GA, Matson R, Cooper A, Ness A, Leary S. Pilot trials in physical activity journals: a review of reporting and editorial policy. Pilot and feasibility studies. 2018;4(1):1-3.
[15] Malterud K. Systematic text condensation: a strategy for qualitative analysis. Scandinavian journal of public health. 2012 Dec;40(8):795-805.
[16] Tong A, et al. Consolidated criteria for reporting qualitative research (COREQ): a 32- item checklist for interviews and focus groups. International journal for quality in health care. 2007 Dec 1;19(6):349-57.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230294

Modelling Information Needs and Sources in a COVID-19 Designated Hospital

Elizabeth M. BORYCKI [a,1], Andre W. KUSHNIRUK[a], Henry OLUKA[a], Simon MINSHALL[a], Kenrick CATO[b], Yalini SENATHIRAJAH[c] and David KAUFMAN[d]
[a]*School of Health Information Science, University of Victoria, Canada*
[b]*School of Nursing, Columbia University, United States of America*
[c]*School of Medicine, University of Pittsburgh, United States of America*
[d]*School of Health Professions, Downstate Health Sciences University, United States of America*
ORCiD ID: Elizabeth Borycki https://orcid.org/0000-0003-0928-8867

Abstract. COVID-19 remains an important focus of study in the field of public health informatics. COVID-19 designated hospitals have played an important role in the management of patients affected by the disease. In this paper we describe our modelling of the needs and sources of information for infectious disease practitioners and hospital administrators used to manage a COVID-19 outbreak. Infectious disease practitioner and hospital administrator stakeholders were interviewed to learn about their information needs and where they obtained their information. Stakeholder interview data were transcribed and coded to extract use case information. The findings indicate that participants used many and varied sources of information in the management of COVID-19. The use of multiple, differing sources of data led to considerable effort. In modelling participants' activities, we identified potential subsystems that could be used as a basis for developing an information system specific to the public health needs of hospitals providing care to COVID-19 patients.

Keywords. Public health informatics, emergency disaster management, information needs, workflow, design, modelling

1. Introduction

The COVID-19 pandemic has had a significant impact on healthcare systems. In the early days of the pandemic, health professionals were surprised by the severity and speed of COVID-19's spread. Many patients quickly deteriorated and required respiratory ventilation. This in turn led to emergency departments being overwhelmed [1]. The spread of the disease led to many health professionals contracting the illness and left healthcare organizations scrambling to identify ways in which patients could be cared for while at the same time protecting caregivers (e.g., physicians, nurses etc.). In response to this crisis, hospitals, healthcare systems, city, state and federal governments worked to monitor the state of disease spread and human as well as material resourcing [1]. In some cities, where the rate of spread of COVID-19 was significant, designated COVID-19 hospitals were created [2]. In this paper we describe our approach to

[1] Corresponding Author: Elizabeth Borycki, Email: emb@uvic.ca

identifying and modelling the information needs of infection control practitioners and hospital administrators responsible for COVID-19 management during the pandemic in a COVID-19 only hospital. Such research could be used to create an integrated informational decision support tool. The work described in this paper is part of a larger study with the objective to develop an evidence-based approach to supporting emergency management disaster response workflow, for health information technology (HIT) mediated access to information and support of work in a COVID-19 designated hospital.

2. Research Objectives

The objectives of this research were to:

1. identify the types of information needs of infection control practitioners and hospital administrators at the height of the COVID-19 pandemic;
2. model the information needs and requirements of infection control practitioners and hospital administrators using UML modeling approaches; and,
3. create a use case diagram showing subsystems and their associated actors and activities to be used in the development of an integrated informational decision support tool.

3. Methods

The researchers interviewed infection control practitioners and hospital administrators to collect data about the types and sources of information that were used to make infection control and administrative decisions surrounding COVID-19 management. Unified Modelling Language (UML) diagrams were used to model information needs, sources of information and workflows both within and external to the hospital.

3.1. Participants and Setting

Individuals were invited to participate in the study via email. All participants were infection control practitioners and hospital administrators in a COVID-19 designated tertiary care facility that serviced a diverse and underserved population in a large metropolitan borough (i.e., population >2 million) of a major North American city. The interviews were conducted by Zoom® and were recorded.

3.2. Procedure

Semi-structured interviews were conducted with each of the participants. The interviews were audio-recorded, transcribed and uploaded into MAXQDA®, a qualitative analysis software tool. The transcripts were coded using a directed content analysis approach [3]. This involved first identifying the individuals (i.e., actors) in COVID-19 response management, followed by directed content analysis of the transcripts with stakeholders to identify themes in the interviews. Interview themes included: (a) information needs (b) activities used to obtain information (and the sources of information), and (c) tasks undertaken to obtain information. Following this, two researchers diagrammed the use cases and subsystems associated with outbreak control and management from the

transcripts. This resulted in the development of a set of UML use case diagrams (i.e., one for each participant). The diagrams were then analyzed to identify key systems and subsystems across actors involved in outbreak response and control. Commonalities across the diagrams were identified to develop a common organizational UML model, involving multiple actors interacting to obtain information needs and to carry out multiple tasks (based on inputs from the interview analysis).

4. Findings

In this section of the paper, we present the demographic and qualitative findings.

4.1. Demographic Data

The data presented in this paper represents a subset of interviews with seven key personnel from larger study, where 53 people were interviewed. Seven individuals (i.e., two males and five females) participated in the study. Participants' administrative work focused on infection control, medical care, nursing care and regulatory compliance (see Table 1).

Table 1. Participant Administrative Work Focus

Administrative Work Focus	Number
Infection Control	3
Medical Care	1
Nursing Care	1
Regulatory Compliance	2

4.2. Qualitative Data

4.2.1. Information Sources

Several disparate sources of infection control data were used by participants. Internal data sources included: (1) electronic health record (EHR) systems, (2) laboratory systems, (3) staffing systems, (4) local policy and procedures, (5) library information systems, (6) contact tracing systems, (7) clinical information systems (includes monitoring of patient clinical status), and (8) outbreak surveillance systems. External sources of data included city, state, and federal websites for COVID-19 reporting. These included: (1) the National Library of Medicine via PubMed®, (2) state government information, (3) federal government information, (4) city government information sites, and (5) public health information sites (e.g., CDC external sources of data were searched and used to support decision making surrounding the spread of the disease, patient treatment/management and protection against COVID-19). Participants had to identify and access multiple sources of internal and external information. This complicated work for individuals and may have caused increased stress and cognitive load.

4.2.2. Information Subsystems

Analysis of the interview data revealed several key themes that can be grouped around the type of activities and subsystems involved in infection control (see Table 2).

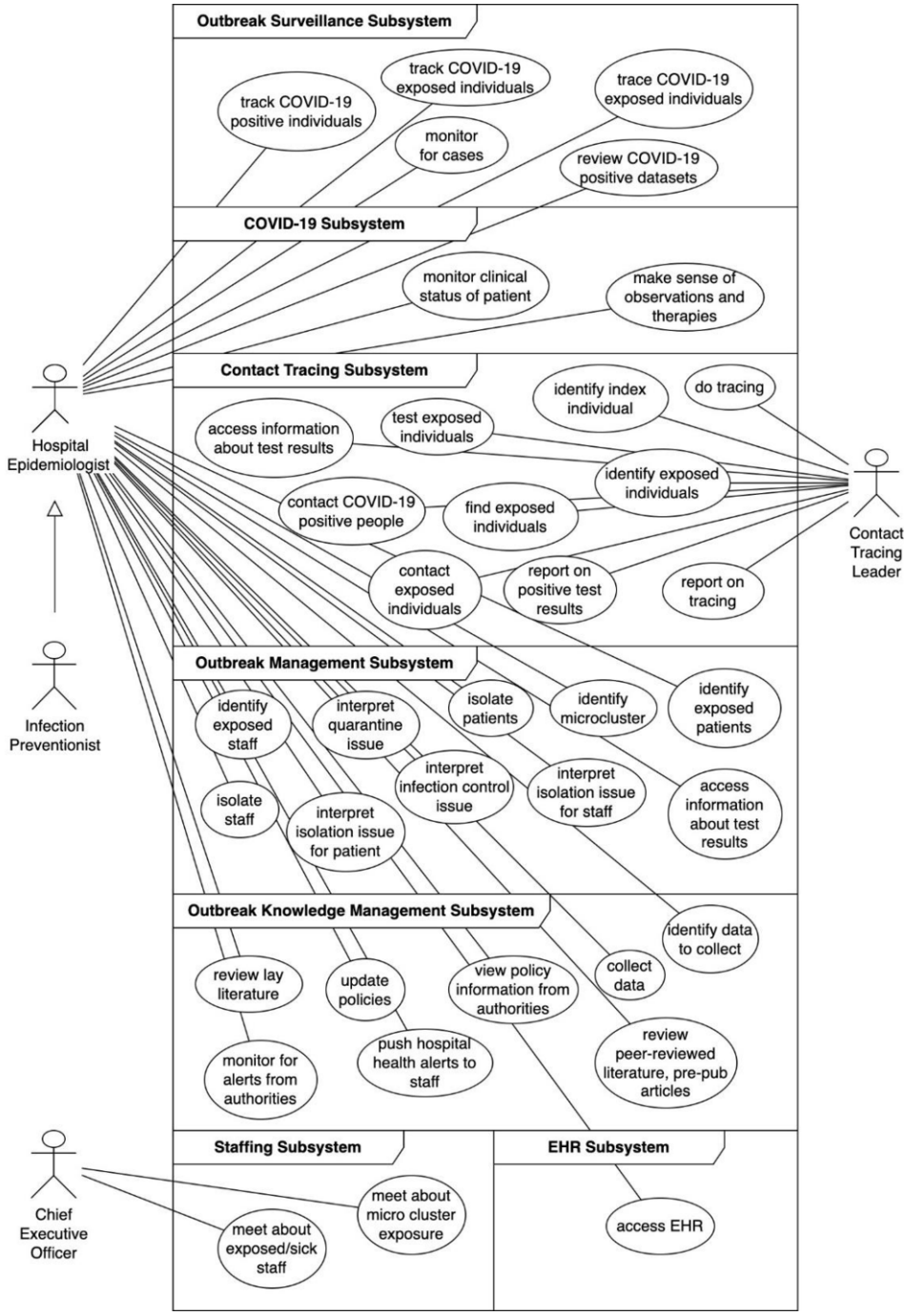


Figure 1: Use Case Diagram

Table 2. Identified Subsystems

Subsystems	Description
Outbreak surveillance subsystem	Monitoring for individuals who have been exposed to COVID-19.
Treatment subsystem	Monitoring of COVID-19 patient clinical status and to determine the efficacy of therapies.
Contact tracing subsystem	Tracking exposed patients and health professionals.
Outbreak management subsystem	Management and monitoring of individuals with COVID-19 so that the infection is not spread.
Outbreak knowledge management subsystem	Monitoring of the latest publications, policies and procedures for managing COVID-19.
Staffing subsystem	Allowing for the management of health professional staff that can be deployed to assist with the management of patients.
EHR subsystem.	Accessing and tracking patient clinical information.

The number of people, who performed specific activities in the hospital, was dependent on resourcing and HIT functionality. This affected the types of information sought by individuals and the amount of work undertaken to obtain and generate information used in hospital decision-making. The issue was exacerbated by a lack of interoperability between systems leading to fragmented workflows that relied on manual work. Participants resorted to creating disparate spreadsheets and dashboards. To complicate matters, regulatory reporting demands increased tenfold, leading to reporting burden. Figure 1 shows an example of a UML use case diagram that details the use cases involved in infection control, including showing the actors involved and their associated activities (based on an analysis of the interview transcripts). In this example, the hospital epidemiologist (who also serves as the infection preventionist) is responsible for many different tasks, drawing from a range of disparate information sources. The diagram illustrates the subsystems involved, and the multiple actors that are part of the workflow.

5. Conclusion: Modelling Information Needs and Sources

There is a need for an integrated information system that allows for easy access to information from disparate sources. This is needed to reduce cognitive burden in hospitals during pandemics. Our research involves analyzing key use cases using UML at a fine-grained level. Using UML modelling, we now have a basis for representing current and to-be activities, and information needs. Such knowledge is foundational to developing HIT, where information fragmentation is present. Information fragmentation leads to individuals looking for information both external and internal to the organization during outbreaks so that essential organizational activities can be maintained. An increase in cognitive load due to accessing so many resources along with information fragmentation leads to difficulty in maintaining organizational level situation awareness. The research in this paper forms the basis for developing information resources and tools to better support information needs during pandemics and reduce cognitive load of actors.

References

[1] WHO. Coronavirus pandemic (COVID19). [cited 2023 March 2] Available from: https://www.who.int
[2] Edwards JA, Breitman I, Kovatch I, Dresner L, Smith TY, Brunicardi FC, Schwartzman A. Lessons learned at a COVID-19 designated hospital. *The American Journal of Surgery*. 2021 Jan 1;221(1):62-4.
[3] Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qualitative Health Research*. 2005 Nov 15(9):1277-88.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230295

Self, at Home, and Digital if Possible, Drivers to Population Health Management

Felix CILLESSEN[a, 1] and Jacob HOFDIJK[b]
[a] *Hospital Rivierenland Tiel, The Netherlands*
[b] *Casemix-CQT Zorg en Gezondheid, Utrecht, The Netherlands*
ORCiD ID: Felix CILLESSEN https://orcid.org/0000-0001-6930-8458,
Jacob HOFDIJK https://orcid.org/0000-0002-4345-3312

Abstract. The Dutch healthcare system is known for its strong emphasis on primary care and a decentralized approach to healthcare delivery. This system will have to adapt in view of the ever-increasing demand and overburdened caregivers, because otherwise it will eventually be insufficient to offer patients adequate care at sustainable costs. The focus must shift from the volume and profitability of all parties involved to a collaborative model for achieving the best outcomes for patients. Rivierenland Hospital in Tiel is preparing for a shift from treating sick patients to promoting the general health and well-being of the population in the region. This "population health" approach aims to maintain the health of all citizens. This transformation to a value-based healthcare system, centered on the needs of patients, requires a complete overhaul of the current systems and its entrenched interests and practices. The regional healthcare transformation requires a digital transformation characterized by several IT implications, such as facilitating patient's access EHR data and sharing information at patient journey level to support the partners involved in the regional care and cure for patients. The hospital is planning to categorize its patients in order to establish an information database. This will help the hospital and its regional partners to identify opportunities for regional comprehensive care solutions as part of their transition plan.

Keywords. ecosystems, patient care planning, care continuity, population health management, integrated health care systems

1. Introduction

The sustainability of the Dutch healthcare system is threatened by a number of factors, including an aging population, the rise of multimorbidity among patients, a shortage of healthcare workers, excessive competition, and rising costs. A shift is necessary from a focus on treating diseases to a focus on maintaining health throughout an individual's lifespan. This will require a change in mindset and a shift towards preventative care and holistic approaches to health and social care. By focusing on self, at home, and digital if possible, we also intend to make better use of the value of data. Telemonitoring services for example, can help patients manage their health remotely and provide valuable data for their healthcare providers, ultimately leading to better outcomes and potentially reducing costs. The Rivierenland hospital, a general hospital in a region with almost

[1]Corresponding Author: Felix Cillessen, Hospital Rivierenland, Department of Medical Informatics, 4002WP Tiel, The Netherlands, Email: Felix.Cillessen@zrt.nl.

200,000 inhabitants, is collaborating with its regional partners to establish a collaboration structure and a new collaborative model of care. With a goal of promoting the idea of "positive health", this effort aims to make a valuable contribution.[1] This will ultimately lead to the healthcare system in the area being more sustainable in the long run.[2]

The Integrated Healthcare Agreement (IZA) in the Netherlands has gained commitment from all stakeholders, resulting in greater regional collaboration and a focus on prevention within the sector.[3] The Ministry of Health is promoting the IZA as the primary policy and allocating funds to support the overhaul. The shift towards a holistic approach to wellness involves reducing the overreliance on medicalization within the healthcare industry. The transition from healthcare to health includes slowing down medicalization. Prompt identification of a patient's health concerns and conducting a thorough and speedy evaluation, if required, is crucial in determining the optimal means of supporting the person.[4] The national healthcare policy is transitioning from a focus on market competition to an emphasis on strengthening regional cooperation. This transition involves the need to realize a digital infrastructure and accelerate the broad adoption of innovations, while also phasing out outdated healthcare practices. As the healthcare system shifts its focus to maintaining health, it must consider in the equation a range of social factors such as housing, living environment, social cohesion, and income policy. In order to effectively address these social issues and bring about meaningful change in the healthcare system, several aspects including funding, stakeholders, governance, and IT systems, must be restructured.

To better understand the patient population and predict the future healthcare needs and costs of specific patient groups, patients can be classified based on their health status, utilization patterns, and demographic characteristics using the Johns Hopkins Adjusted Clinical Groups (ACGs) system. The ACG system has been found to be a useful tool for stratifying Dutch healthcare populations.[5] The value of the data can also be improved by assessing the Kaiser Permanente risk model as the foundation for discussions on shifting to a model of care that emphasizes multimorbidity team work and a digitally-based regional care model.

The EU4Health is a European Union program that also aims to improve healthcare and foster innovation. It prioritizes areas like promoting health and preventing illness, implementing digital technology, and developing the healthcare workforce.[6] The objectives of the EU4Health program are considered part of the Tiel approach.

2. Methods

Various studies have been carried out in which the research results have been used as input for drawing up a transition plan that should be ready by mid-2023.

For the past 4 years, a collaborative effort between diverse care providers and the social sector has been dedicated towards developing novel approaches to provide optimal care to senior citizens right where they need it. The goal is to ensure that senior citizens are well-equipped for their later years and can enjoy good health, while residing in a comfortable and familiar setting, with access to medical assistance if needed. In 2020, a non-profit care consultancy (MURA) carried out a research study on the present and upcoming conditions of care and elderly care.

In 2020, the largest health insurer in the region (MENZIS) published a regional vision for 2030 as an extra impulse for regional cooperation and prevention. The vision was created in collaboration with local healthcare organizations, general practitioners,

municipalities and patient associations and shows how supply and demand for health maintenance and patient care will develop over the next 10 years.

The OECD (Organization for Economic Co-operation and Development) 'Health at a Glance 2022' report to compare key indicators for population health and health system performance across OECD member countries and key emerging economies provide lead principles for the transition.[7]

In the year 2020, the formation of a committee dedicated to innovation marked the start of a mission to advance digital technology and information systems, with a focus on making healthcare processes "future-proof" through technology rather than solely prioritizing the technology itself. Subsequently, in 2021, the initiation of an EHR optimization process ensued, emphasizing consistent registration procedures across all specialties to ensure uniform online access for patients.

Finally, an administrative consultation has been organized every quarter from 2020. All parties involved in providing, organizing, or paying for care are represented in this platform, including the directors of these care providers. It plays a crucial role developing ideas and finding solutions for addressing the care challenges for the Rivierenland area.

3. Results

MURA expects the number of elderly aged 75+ to increase from 8% in 2019 to 17% in 2050. In June 2020, the Regional Picture of the Elderly in Rivierenland was presented to professionals and directors which resulted in three recommendations:

1. **Prevention:** through a healthy lifestyle: Encourage seniors to maintain a healthy lifestyle, with an emphasis on physical activity and nutrition. Maintaining social connections is also important and can be a motivator for physical activity. Positive health and solution-focused approaches are important for prevention: How can we motivate people? What brings happiness and pleasure? What works well and what can we learn from it?
2. **Appropriate housing**: Seek out and provide appropriate housing that meets the needs and preferences of seniors. This includes not just adapted homes, but also considering the physical and social living environment and whether it invites movement and social interaction.
3. **Empowering seniors:** to have more control over aging: Focus on helping seniors maintain or increase control over their health and life as they age. This includes helping seniors become aware of their future and supporting them in making choices to maintain or increase their control, such as by having early conversations about their preferences and needs as they age.

The MENZIS study identifies two important developments:

1. **Development of care demand**. The population in the region is growing by 2% until 2030. There has been a 55% increase in the number of residents with dementia in the region due to the aging of the population. (48% at a national level). The number of chronically ill people is growing more than the average in the Netherlands. Joint wear and tear and cardiovascular diseases have experienced the most significant growth. Mental illnesses (mood disorders and anxiety) are growing only slightly in the coming years. The proportion of residents with overweight is increasing (from 52% to 60%), while the proportion of smokers is decreasing (from 20% to 15%).
2. **Development of care supply**. The number of care providers is expected to decrease by about 5% until 2030. The shortage of specialists in geriatric medicine is expected

to continue to increase. In mental health care (GGZ), the vacancy rate of all professions has strongly increased in recent years. A further increase in the shortage is also expected here. In terms of absolute numbers, the shortage is largest in nursing. The majority of the projected 80,000 care staff shortages in the Netherlands for the year 2022 can be attributed to this factor. Sick leave in the care sector has been increasing since 2014. Absence is almost a third higher than the average for employees in the Netherlands. The turnover in the care sector has also been increasing in recent years, standing at almost 16%. Informal care (the potential for informal caregivers) will decrease in the Rivierenland region by almost a third until 2030, but will remain more favorable than the average in the Netherlands.

Important conclusions from the OECD report are that lifestyle choices that are harmful to one's health, as well as poor environmental conditions, are ongoing issues that negatively impact the quality of life, reduce lifespan, and make populations less resilient to health crises. Additionally, relatively little money is spent on disease prevention, which makes up only about 3% of overall health spending on average.

As of today, patients can choose from three methods to access their records, namely the patient portal, Blue Button download, or a personal health environment. Additionally, two pilot telemonitoring programs have been launched to monitor patients with chronic obstructive pulmonary disease (COPD) and heart failure (HF). A dedicated service desk has also been established to offer guidance for digital care. Furthermore, care providers who are registered can now access a patient's hospital EHR, provided the patient has granted permission.

A plan for transformation is presently in the works, aimed at providing guidance to establish a sustainable healthcare system in the area. To further shape the transformation, various changes are being implemented. These transformations are organized under three pillars: 1) Care in motion, 2) Innovation and digitization, and 3) Patient in control.

4. Discussion

Self, at home, and digital if possible, will have to prove to be a critical success factor important in sustaining the effectiveness, accessibility, and affordability of healthcare.

The advancements made towards the healthcare system in the Rivierenland area will lead to an interconnected network that prioritizes the physical and mental health of the community, offers tailor-made treatment to patients, streamlines the overall process of healthcare providers, and ultimately achieves better health results while simultaneously cutting down expenses. To shift towards a sustainable, value-based health care system, it is necessary to foster a distinct form of collaboration involving bundled payments, which are intricate and comparatively novel.[8] This transformation brings together various stakeholders to address common challenges and inefficiencies. Much can be learned from the Greater Manchester project, where trust was the fundamental underlying issue for transformation of the health care model.[9]

By stratifying patients, we can gain a deeper understanding of how costs are distributed, the degree of multimorbidity, and potential inefficiencies within the delivery of care. It will unlock the potential of the Epital Care Model (ECM) for the introduction of digital care.[10] The operationalization of the concepts of Self, Home, and Digital through the ECM model make it a valuable tool for driving the IZA transition. By utilizing patient-centered data, we have the potential to identify and minimize wasted

resources, while also enhancing efficiency. The first small, yet significant, steps towards patient-centered care will be taken to achieve the required transition to IZA standards.

Currently, electronic health records often only include a limited list of issues addressed at the institutional level. True person-centered care requires a more comprehensive and holistic approach that results in a comprehensive overarching care plan of the patient journeys. There is a lack of awareness among healthcare providers regarding patients' ability to access their data, the type of data that is viewable, how the data is organized, and the fact that data may have different names in various systems.

5. Conclusion

Care will become unaffordable if we continue as we are now. Relatively little money is spend on maintaining health, while the greatest profit seems to be achievable there. Success is heavily reliant on establishing a regional governance structure and embracing digital transformation, which in turn requires strong collaboration with regional partners. The transformation towards appropriate care requires innovations in technology and digital solutions to make healthcare more efficient and effective. One way to enhance healthcare and advance digital/hybrid care models is by leveraging the insights gleaned from data. To achieve this, it's essential to prioritize improving the digital competencies of staff since their proficiency is a key determinant of success.

References

[1] Huber M, van Vliet M, Giezenberg M, Winkens B, Heerkens Y, Dagnelie PC, et al. Towards a 'patient-centred' operationalisation of the new dynamic concept of health: a mixed methods study. BMJ Open. 2016 Jan;6(1):e010091.doi:10.1136/bmjopen-2015-010091.

[2] Hofdijk J, Cillessen F, Asin J. The Tiel Shift Towards Person Centered Care. Stud Health Technol Inform. 2022 May;294:699-700. doi:10.3233/SHTI220559.

[3] "Integraal zorgakkoord." Dutch government. https://www.rijksoverheid.nl/ministeries/ministerie-van-volksgezondheid-welzijn-en-sport/documenten/rapporten/2022/09/16/integraal-zorgakkoord-samen-werken-aan-gezonde-zorg (accessed Jan 2023).

[4] Cillessen F, Hofdijk J. Transition Requirements from Problem List to an Overarching Care Plan for the Support of Person-Centered Care. Stud Health Technol Inform. 2020 Jun;272:292-5. doi: 10.3233/SHTI200552.

[5] Girwar S-AM, Fiocco M, Sutch SP, Numans ME, Bruijnzeels MA. Assessment of the Adjusted Clinical Groups system in Dutch primary care using electronic health records: a retrospective cross-sectional study. BMC Health Services Research. 2021;21(1):217. doi: 10.1186/s12913-021-06222-9.

[6] "EU4Health – a vision for a safer and healthier European Union 2021-2027." https://ec.europa.eu/health/funding/eu4health_en (accessed Jan 2023).

[7] OECD/European Union (2022), Health at a Glance: Europe 2022: State of Health in the EU Cycle, OECD Publishing, Paris, https://doi.org/10.1787/507433b0-en.

[8] Steenhuis S, Struijs J, Koolman X, Ket J, VAN DER Hijden E. Unraveling the Complexity in the Design and Implementation of Bundled Payments: A Scoping Review of Key Elements From a Payer's Perspective. Milbank Q. 2020 Mar;98(1):197-222. doi:10.1111/1468-0009.12438.

[9] Elvey R, Bailey S, Checkland K, McBride A, Parkin S, Rothwell K, et al. Implementing new care models: learning from the Greater Manchester demonstrator pilot experience. BMC Fam Pract. 2018 Jun;19(1):89. doi: 10.1186/s12875-018-0773-y.

[10] Phanareth K, Vingtoft S, Christensen AS, Nielsen JS, Svenstrup J, Berntsen GK, et al. The Epital Care Model: A New Person-Centered Model of Technology-Enabled Integrated Care for People With Long Term Conditions. JMIR Res Protoc. 2017 Jan;6(1):e6. doi: 10.2196/resprot.6506.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230296

An Adaptive Digital Intelligence System to Support Infodemic Management: The WHO EARS Platform

Becky WHITE[a], Ivan CABALERO[b], Tim NGUYEN[a], Sylvie BRIAND[a], Agnese PASTORINO[a] and Tina D PURNAT[a,1]
[a] *Department of Epidemic and Pandemic Preparedness and Prevention, WHO Health Emergencies Programme, World Health Organization, Geneva, Switzerland*
[b] *Citibeats, Barcelona, Spain*

Abstract. The WHO Early AI-Supported Response with Social Listening (EARS) platform was developed to help inform infodemic response during the COVID-19 pandemic. There was continual monitoring and evaluation of the platform and feedback from end-users was sought on a continual basis. Iterations were made to the platform in response to user needs, including the introduction of new languages and countries, and additional features to better enable more fine-grained and rapid analysis and reporting. The platform demonstrates how a scalable, adaptable system can be iterated upon to continue to support those working in emergency preparedness and response.

Keywords. Social listening, infodemic, AI, COVID-19, pandemic response, pandemic preparedness, social media monitoring

1. Introduction

The infodemic during the COVID-19 pandemic has impacted poorly on public health [1]. The WHO EARS platform was developed to enable infodemic managers to understand COVID-19-related narratives online. EARS uses Artificial Intelligence (AI) and machine learning to categorize publicly available digital and social media data to a public health taxonomy. Understanding the concerns, questions, information voids and narratives among citizens can help to inform pandemic and infodemic response [2]. The platform was launched in December 2020, initially covering 20 countries and four languages, and has analyzed over 86 million posts to October 2022.

2. Methods

Regular internal processes to understand the needs of those using the EARS system were implemented. Evaluation was conducted during training sessions, as written reviews, and through user consultations. Technical review was conducted each month with the development team to review global trends, the public health taxonomy, data sources and

[1] Corresponding author: Tina Purnat, E-mail: purnattt@who.int.

data volume. The following were identified from end-users: need for more fine-grained exploration of COVID-19 vaccine narratives; features to better support rapid analysis; more streamlined reporting features; and broader coverage of countries and languages.

3. Results

The EARS platform underwent iterations in response to user needs. Several new features were added to the platform to enable faster identification of insights and reporting. A 'stories' header on the main analysis page summarizes key rising narratives, alerting users to gender, category and country indications. A new sophisticated 'social indictors' panel helps users to identify social change in narratives across categories such as distrust, civic unrest, or polarization. An added reporting feature allows users to create, store and collaborate on producing reports, including options for automation. An additional five languages and 10 countries were added, and iterations made to the taxonomy. To enable more informed narratives about the COVID-19 vaccine roll-out, a new dashboard interface was added. This involved the development of a new vaccine-specific taxonomy comprising of 21 categories, for 15 priority countries and 10 languages.

4. Discussion

The innovative EARS platform is the first we are aware of to offer free, real-time access to data, categorized to a public health taxonomy, to enable rapid social understanding. Much social listening research is from high income countries [3] and EARS platform filters enable prioritization of narratives by gender, country or language. The needs of end-users evolved throughout the pandemic and the rounds of evaluation and iteration enabled the platform to remain relevant and useful. The adaptable and scalable use of AI technology to inform response and adapt to user needs is an example of how technology can support the changing needs of infodemic managers during a health emergency.

5. Conclusion

Digital social listening platforms can provide useful data, which when combined with other sources can produce actionable insights to guide response. Remaining agile and adaptable during a health emergency is important, the work done with the WHO EARS platform highlights how digital intelligence gathering can pivot to need changing needs.

References

[1] Islam MS, Sarkar T, Khan SH, Mostofa Kamal A-H, Hasan SMM, Kabir A, Yeasmin D, Islam MA, Amin Chowdhury KI, Anwar KS, Chughtai AA, Seale H. COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. The Am J Trop Med. 2020;103(4):1621-9.

[2] Purnat, TD, Nguyen, T, Yau, B, White, B, Cecchini, S, Samuel, R, Hess, S, Bezbaruah, S, Briand, S. Delivering actionable infodemic insights and recommendations for the COVID-19 pandemic response. Weekly Epidemiological Record. 2022 Jul 8;97(27):313-24.

[3] Skafle I, Nordahl-Hansen A, Quintana DS, Wynn R, Gabarron E. Misinformation About COVID-19 Vaccines on Social Media: Rapid Review. J Med Internet Res. 2022;24(8):e37367. doi:10.2196/37367.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230297

Narrative Trends over the COVID-19 Pandemic: Digital Social Listening to Inform WHO Infodemic Management

Becky WHITE[a], Atsuyoshi ISHIZUMI[a], Brian YAU[a], Amy WRIGHT[b], Lucy LAVERY[b], Tim NGUYEN[a], Sylvie BRIAND[a], Tim ZECCHIN[b] and Tina D PURNAT[a,1]

[a] *Department of Epidemic and Pandemic Preparedness and Prevention, WHO Health Emergencies Progamme, World Health Organization, Geneva, Switzerland*

[b] *Marble Global, London, UK*

Abstract. The COVID-19 infodemic is an overwhelming amount of information that has challenged pandemic communication and epidemic response. WHO has produced weekly infodemic insights reports to identify questions, concerns, information voids expressed and experienced by people online. Publicly available data was collected and categorized to a public health taxonomy to enable thematic analysis. Analysis showed three key periods of narrative volume peaks. Understanding how conversations change over time can help inform future infodemic preparedness and prevention planning.

Keywords. Social listening, infodemic, COVID-19, pandemic response, pandemic preparedness, social media, taxonomy

1. Introduction

The infodemic accompanying the COVID-19 pandemic has led to an overwhelming about of information, particularly on social and other media [1]. The infodemic can lead to confusion, reduced trust in health authorities and an increase in risk-taking behavior. As part of understanding and managing the infodemic, WHO has worked with partners to produce weekly digital infodemic intelligence reports since February 2020. This communication outlines data collection and narrative peaks between March 2020 and October 2022.

2. Methods

Publicly available social and news media data is collected from Meltwater and CrowdTangle on a weekly basis in English, French and Spanish. These data are categorized to a public health taxonomy which has 5 overarching categories (the cause of the virus, the illness, the treatment, the interventions and perceptions on information) and 42 sub-categories [1]. As well as analysis of data volume, data analysis reporting

[1] Corresponding author: Tina Purnat, E-mail: purnattt@who.int.

also includes information voids, top questions, and data velocity to identify rising topics. Data within categories is analyzed qualitatively to identify narrative trends.

3. Results

Between March 2020 and October 2022, over 1.88 billion global social media posts were identified that mentioned COVID-19. Between January 2021 and October 2022, 381 million posts discussed COVID-19 vaccines, making it the most prominent topic of conversation. There were three key periods associated with high data volumes. A peak in March – April 2020 was conversation driven by misinformation and conspiracy theories about the origin of COVID-19, discussion about the impact of socio-economic factors and underlying conditions on COVID-19 reported deaths, and discussion about efficacy of unproven treatments. In October 2020, a large spike in the 'Personal Measures' category was driven by conversations about the use of protective equipment, transmission risk, and the severity of COVID-19. Finally, between December 2021 and January 2022, Omicron-related narratives were most discussed. In addition there were discussions about travel restrictions, debate about the severity of Omicron, and if this variant indicated we were moving towards an endemic phase. There were also narratives regarding vaccination, which brand of vaccine was best for a booster dose, as well as frustration at the spread of vaccine questioning misinformation.

4. Discussion

Using a taxonomy to digital social listening has enabled better understanding of global narratives, concerns and information voids throughout the pandemic. Analysis of narratives trends and volume peaks has shown how conversations change over time as knowledge about COVID-19 increased, as the epidemiology of the disease evolved, and as vaccines and public health and social measures were used. Understanding these trends offer lessons for future preparedness and prevention planning.

5. Conclusion

Digital social listening can provide useful insight into how citizens are experiencing an emergency, and their information and other needs. Using a public health taxonomy has enabled infodemic managers to triangulate with other data sources and translate insights faster into recommendations.

References

[1] Tangcharoensathien V, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res. 2020;22(6):e19659. doi:10.2196/19659.

[2] Purnat, TD, Nguyen, T, Yau, B, White, B, Cecchini, S, Samuel, R, Hess, S, Bezbaruah, S, Briand, S. Delivering actionable infodemic insights and recommendations for the COVID-19 pandemic response. Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire. 2022 Jul 8;97(27):313-24.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230298

'Smart' Buffalo Weight Estimation via Digital Technologies: Experiences from South Italy

Oscar TAMBURIS[a,b 1], Roberta MATERA[b], Angela SALZANO[b], Antonio CALAMO[b], Domenico VISTOCCO[c] and Gianluca NEGLIA[b]
[a] *Institute of Biostructures and Bioimaging, National Research Council, Naples, Italy*
[b] *Dept. of Veterinary Medicine and Animal Productions, University "Federico II", Naples, Italy*
[c] *Dept. of Political Sciences, Communications and International, University "Federico II", Naples, Italy*

Abstract. The present work aims at describing a viable "protocol" for unobtrusive direct/indirect monitoring of biometric parameters for the estimation of body conditions on Mediterranean Buffalo populations, using low-cost automated systems i.e., smart cameras endowed with depth perception capabilities.

Keywords. Public Health Informatics, Precision Livestock Farming, Mediterranean buffalo, Smart cameras, 3D/2D image analysis

1. Introduction

Precision Livestock Farming (PLF) is reported as a declination of Public Health Informatics [1,2] focused on the application of process engineering principles and techniques to livestock farming in order to an automatic supervision, modelling, and management of animal production. About one ninth of the global cattle population is composed by an essential domestic bovid, the so-called water (or river) buffalo (*Bubalus bubalis*) [3]. In South Italy the whole Buffalo–related dairy production and supply chain represents a leading sector of the entire Agri–food arena, and the awareness is currently increasing about the importance of a data-driven livestock management to cope with the complexity of decision-making processes. the present study a set of Machine Learning–based algorithms were deployed on the set of measurements obtained by a combination of three low-cost smart cameras endowed with depth perception capabilities, and compared with traditionally hand-performed measurements, on Mediterranean Buffalo calves from the birth to their complete weaning.

2. Materials and Methods

The trial lasted 90 days (13 weeks) and was carried out in the period June-September 2022. A longitudinal observational study was conducted on 30 female buffalo calves, from their birth up to the weaning phase. To predict body weight, for each of them every

[1] Corresponding Author: Oscar Tamburis, IBB CNR, Via Tommaso de Amicis 95, 80145 Naples, Italy. E-mail: oscar.tamburis@ibb.cnr.it

week the following three biometric measurements were taken: (i) height at the withers (WH) i.e., vertical distance between the withers (highest point of the back, between the neck and shoulder blades) and the ground; (ii) body length (BL) i.e., oblique distance between the tip of the buttock (apophysis of the ischium) and the tip of the shoulder (shoulder joint); (iii) chest girth (CG) i.e., minimum value measured just behind the shoulders. Two Intel ® RealSenseTM cameras (Depth camera D415 and LiDAR camera L515) along with a RICOH ® WG–60 photo camera, were used for measuring WH and BL. Three different models – Multiple Linear and Polynomial Regression (MLR/MPR), and Artificial Neural Network (ANN) – were implemented (and compared) to predict the progression of calves' body weight during the trial, starting from the body measurements taken [4].

3. Results

2195 single evaluations of WH, BL, CG, and body weight were performed. Mean Squared Error (MSE), Root MSE (RMSE), and Pearson's R^2 goodness-of-fit criteria were used to evaluate the performance of the model. For what concerns WH the lowest values for both MSE and RMSE appear to be those related to L515 (AVG_{MSE}=11,41; AVG_{RMSE}=3,02). The result is only apparently better, given the smaller number of measurements taken with this tool, thus returning more reliable values than those from D415 (AVG_{MSE}=15,03; AVG_{RMSE}=3,72). BL features instead an overall better trend of D415 (AVG_{MSE}=29,78; AVG_{RMSE}=4,99). Besides the manual measurements, the best prediction is the one performed by means of ANN, starting from the measurements taken with the Stereo camera D415.

4. Discussion and Conclusions

The experimentation conducted aimed at proposing, formalizing, and testing a viable "protocol" for unobtrusive direct/indirect monitoring of biometric parameters for the estimation of body conditions on Mediterranean Buffalo populations, in order to: (i) figure out the most timely measurement tools to be used to correctly estimate and predict the weight trajectories; and (ii) verify the suitability of the adoption of an electronic measurement system to achieve best practices in Precision Livestock Breeding, thus showcasing that the continuous improving of the Smart Farming sector is meant to widen Public Health Informatics policies and strategies [5].

References

[1] Magnuson JA, Dixon BE, editors. Public health informatics and information systems. Springer Nature; 2020 Jul 17.
[2] BUFF4L.0: Veterinary and Engineering Sciences at the crossroads in the Industry 4.0 age. In 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI) 2021 Sep 6 (pp. 137-141). IEEE.
[3] Matera R, Di Vuolo G, et al. Relationship among Milk Conductivity, Production Traits, and Somatic Cell Score in the Italian Mediterranean Buffalo. Animals. 2022 Aug 29;12(17):2225.
[4] Song X, Bokkers EA, et al. Automated body condition scoring of dairy cows using 3-dimensional feature extraction from multiple body regions. Journal of dairy science. 2019 May 1;102(5):4294-308.
[5] Benis A, Tamburis O, Chronaki C, Moen A. One digital health: a unified framework for future health ecosystems. Journal of Medical Internet Research. 2021 Feb 5;23(2):e22189.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230299

Identification of Subphenotypes of Opioid Use Disorder Using Unsupervised Machine Learning

Fatemeh SHAH-MOHAMMADI[1] and Joseph FINKELSTEIN
Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract. This paper aimed to detect the latent clusters of patients with opioid use disorder and to identify the risk factors affecting drug misuse using unsupervised machine learning. The cluster with the highest proportion of successful treatment outcomes was characterized by the highest percentage of employment rate at admission and discharge, the highest percentage of patients who also recovered from alcohol and other drug co-use, and the highest proportion of patients who recovered from untreated health issues. Longer participation in opioid treatment programs was associated with the highest proportion of treatment success.

Keywords. Opioid use disorder, Machine learning, Subphenotyping.

1. Introduction

Optimal opioid use disorder (OUD) treatment in opioid treatment programs (OTP) should be tailored to individual patient profiles [1]. The goal of this study was the identification of subphenotypes of patients with OUD using heterogenous data sources generated by OTP clinics and mapping the uncovered latent classes to OTP outcomes.

2. Methods

The study cohort comprised patients with OUD undergoing treatment at OTP clinics at the Mount Sinai Health System (MSHS) in New York City. The study dataset was generated by aggregation, and harmonization of patient information from the New York State Office of Addiction Service and Supports (OASAS), OTP clinic software for patient management (AVATAR), and MSHS electronic health records (Epic). The analytical dataset comprised patients' socio-demographic information, clinician notes, drug urine toxicology screens, employment status, admission, transfer, and discharge records, and state forms. The discharge form contained information on the patient's drug and alcohol recovery status, and data about the medical and legal issues. OTP outcome was defined based on discharge summary, which included discharge status as follows: "completed treatment: all treatment goals met", "completed treatment: half or more goals met", "treatment not complete: some goals met," and "treatment not complete: no goals

[1] Corresponding Author: Fatemeh Shah-Mohammadi, Icahn School of Medicine at Mount Sinai, 1770 Madison Ave, 2nd Fl, New York, NY, USA, 10035; E-mail: fateme.sh.mohammadi@gmail.com.

met". The patients with the first two discharge statuses were considered as patients with successful treatment, and patients with the last two statuses were labeled as failed treatment cases. We used the factor analysis for mixed data (FAMD) method for preprocessing [2].

3. Results and Discussion

Three clusters have been identified. There was a significantly larger number of male patients compared to females. Most patients in all clusters were 45 years old and older. All patients who succeeded in treatment belonged to cluster # 2. Compared to the patients in clusters # 0 and 1, patients in cluster # 2 had the highest employment rate at admission and discharge (18.47% and 19.79%, respectively). The percentage of patients who recovered from alcohol and drug abuse (43.47% and 86.97%, respectively) was significantly higher in cluster # 2. The same scenario was valid in meeting the legal and medical goals for patients in cluster # 2. Compared to other clusters, the rate of patients who adhered to the treatment for over 2 years was almost twice as high in cluster # 2. Since the lower percentage of employment at admission and discharge, achievement of alcohol, drug, medical and legal goals belonged to the clusters with the largest percentage of failed patients (i.e., clusters # 0 and 1), patients in these clusters need more rigorous treatment, social support, and dedicated personal attention. This finding conveys the fact that OTP patients with alcohol co-use are at risk of worse treatment outcomes. Due to the increase in employment rate from admission to discharge among the patients in the cluster with the highest successful treatment rate, it can be assumed that employment acts as a facilitator for adhering to treatment for opioid use disorders and can result in better treatment outcomes. This finding is aligned with many studies claiming that unemployment is a significant risk factor for substance use and the subsequent development of substance use disorders. Due to the highest rate of patients with more than 2 years of participation in OTP in cluster #2, it can be concluded that the longest enrollment in OTP is associated with the highest successful treatment rate.

4. Conclusion

Three clusters have been identified for patients who participated in OTPs. The largest percentage of successfully treated patients belonged to the second cluster, with patients mainly employed. This conveys that unemployment acts as a risk factor for the treatment failure of opioid use disorders. This cluster also has the highest percentage of patients who recovered from untreated health issues from alcohol and other drug co-use. The longest adherence to the treatment is associated with the highest successful treatment rate. Subphenotyping of patients with OUD may help personalize their treatment.

References

[1] McCarthy M. Opioid prescribing rates in US vary widely between states, CDC reports. BMJ. 2014 Jul 4;349:g4424.

[2] Mori Y, Kuroda M, Makino N Multiple Correspondence Analysis. In: Nonlinear Principal Component Analysis and Its Applications. Springer Briefs in Statistics. Springer, Singapore.2016.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230300

Converting HL7 CDA Based Nationwide Austrian Medication Data to OMOP CDM

Florian KATSCH[a,b,1], Rada HUSSEIN[a],
Raffael KORNTHEUER[b] and Georg DUFTSCHMID[b]
[a] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[b] *Center for Medical Data Science, Medical University of Vienna, Vienna, Austria*

Abstract. Austria's national Electronic Health Record (EHR) system holds information on medication prescriptions and dispenses in highly structured HL7 Clinical Document Architecture (CDA) documents. Making these data accessible for research is desirable due to their volume and completeness. This work describes our approach of transforming the HL7 CDA data into Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) and highlights a key challenge, namely mapping the Austrian drug terminology to OMOP standard concepts.

Keywords. Electronic Health Record, Electronic Prescription, HL7 CDA, OMOP CDM, Health Information Interoperability

1. Introduction

Since 2019 Austria has been operating a national health information exchange (HIE) system for medication data. All prescriptions from outpatient care providers and all dispenses in pharmacies are included in structured form, based on the HL7 Clinical Document Architecture (CDA) standard. Currently, about 97% of the 9 million Austrian inhabitants participate in the HIE system. Roughly 10 million medication prescriptions and dispenses are registered to the system monthly [1].
The OMOP CDM is a well-established standardized representation of clinical data that is broadly used in inter-institutional research projects. In this work we describe our project for transforming Austrian's CDA-formatted medication data to the OMOP CDM, supported by the European Health Data and Evidence Network (EHDEN).

2. Methods

In the HL7 CDA data model, prescriptions and dispenses are represented via classes *SubstanceAdministration* and *Supply*. Both reference class *ManufacturedProduct*, which holds the drug to be consumed. The CDA data model is customized to the settings of concrete HIE use cases via CDA templates. The Austrian CDA medication templates [2] are conformant to the corresponding IHE Patient Care Coordination templates [3] to achieve international interoperability. In contrast to [4], who used absolute XPaths to

[1] Corresponding Author: Florian Katsch, E-mail: florian.katsch@dhp.lbg.ac.at.

refer to CDA content, we apply relative XPaths based on template identifiers within our extract-transform-load (ETL) specification. Besides several other CDM tables, *DRUG_EXPOSURE* represents the main target of the CDA medication data. As CDA medication classes are highly structured, 17 from 23 *DRUG_EXPOSURE* attributes can be fed. Attributes *drug_concept_id* and *route_concept_id* are supplied from data coded in Austrian-specific terminologies and thus need to be mapped to CDM standard concepts. In our ETL activities we are supported by an EHDEN certified SME.

3. Results

We have completed the EHDEN ETL definition document that specifies the mappings. Currently, we are working on the ETL implementation by means of a Python project. After setting up prerequisite components (such as the database connectors and test infrastructure) we continued with supplying the commonly used CDM tables (e.g. *PERSON*, *CARE_SITE*). Meanwhile we have started integrating the clinical fact tables.

4. Discussion and Conclusion

A key challenge of our work is to map Austrian drug codes (PZN) to the CDM standard code system RxNorm. Each PZN is associated with at least one ATC code and OMOP features a mapping from ATC to RxNorm. However, ATC codes only allow mapping to the rather coarse RxNorm *Ingredient* level. Thus we lose information about drug strength and dose form. Furthermore, due to different possible indications for a drug administration, a PZN may have multiple ATC codes assigned, which introduces additional ambiguities. A solution could be to make use of the "Austria-Codex" drug ontology that includes, amongst others, active substances, drug strength, and dose form. As these attributes are also available in OMOP for the more granular levels of RxNorm, an ontology alignment seems possible. We plan to explore this alignment with clinical domain experts by focusing on a limited set of diabetes-specific drugs that are particularly relevant in a current research project. Our work will pave the way for Austrian medication data to be usable for research on a national scale. Since our CDA medication templates are conformant to the IHE Patient Care Coordination medication templates, our work has potential for international reuse.

References

[1] ELGA GmbH. ELGA: Zahlen-Daten-Fakten [Internet]. 2022 [cited 2023 Mar 2]. Available from: https://www.elga.gv.at/elga/zahlen-daten-fakten/

[2] ELGA GmbH. Implementierungsleitfaden E-Medikation (Version_2.06) [Internet]. 2015 [cited 2023 Mar 2]. Available from: https://www.gesundheit.gv.at/dam/jcr:5abf0398-6785-433b-a3d4-d0e0972c6b03/Implementierungsleitfaden_E-Medikation_(Version_2.06).pdf

[3] IHE. IHE Patient Care Coordination (PCC) 5 Technical Framework [Internet]. 2021 [cited 2023 Mar 2]. Available from: https://www.ihe.net/uploadedFiles/Documents/PCC/IHE_PCC_TF_Vol2.pdf

[4] Ji H, Kim S, Yi S, Hwang H, Kim J-W, Yoo S. Converting clinical document architecture documents to the common data model for incorporating health information exchange data in observational health studies: CDA to CDM. J Biomed Inform. 2020 Jul;107:103459.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230301

Prognostic Factors for Covid-19 on Admission Profile and Air Pollutants

Kikue SATO[a,1], Taiki FURUKAWA[a], Satoshi YAMASHITA[a], Daisuke KOBAYASHI[b], Shintaro OYAMA[c] and Yoshimune SHIRATORI[a]
[a] *Nagoya University Hospital Medical IT Center, Nagoya University, Nagoya, Japan*
[b] *Graduate School Medicine/School of Public Health Department of Health Economics and Quality Management, Kyoto University, Kyoto, Japan*
[c] *Innovative Research Center for Preventive Medical Engineering, Nagoya University, Nagoya, Japan*

Abstract. It has been reported that the severity and lethality of Covid-19 are associated with coexisting underlying diseases (hypertension, diabetes, etc.) and cardiovascular diseases (coronary artery disease, atrial fibrillation, heart failure, etc.) that increase with age, but environmental exposure such as air pollutants may also be a risk factor for mortality. In this study, we investigated patient characteristics at admission and prognostic factors of air pollutants in Covid-19 patients using a machine learning (random forest) prediction model. Age, Photochemical oxidant concentration one month prior to admission, and level of care required were shown to be highly important for the characteristics, while the cumulative concentrations of air pollutants SPM, NO2, and PM2.5 one year prior to admission were the most important characteristics for patients aged 65 years and older, suggesting the influence of long-term exposure.

Keywords. AirPollutionExposure, Covid-19, MachineLearning, Prognostic Factors

1. Introduction and Methods

The health effects of air pollutants on deaths from cardiovascular and respiratory diseases have become clear in many countries [1], and it has been reported that air pollution and other environmental exposures may be included as risk factors for mortality in Covid-19 as well [2]. If prognosis can be predicted based on environmental exposures at the individual level, based on patient characteristics at the time of admission and residential information, it will contribute to the evaluation of causal relationships between health effects and mortality.

DPC (Diagnosis Procedure Combination) is a patient classification method developed in Japan for inpatients in the acute phase of illness, and the nationwide uniform electronic DPC data includes clinical information on patients, information on medical procedures used for patient classification. Using the DPC data, this study included 4,071 cases of Covid-19 (ICD-10: U07.1 or U07.2) who were hospitalized at medical institutions in Aichi Prefecture from April 2020 to April 2021.

Air pollutants covered were nitrogen dioxide (NO2), photochemical oxidants (Ox), suspended particulate matter (SPM), and fine particulate matter (PM2.5). Air pollution

[1] Corresponding Author: Kikue Sato, E-mail: satokiku@med.nagoya-u.ac.jp.

concentrations are calculated using monthly averages from each air pollution monitoring station located in designated areas by local governments, and are compared with monitoring stations located in the vicinity of each patient's residence. Cumulative exposures for the previous month and year are obtained from the date of hospitalization for Covid-19. Air pollution constant monitoring data are available at the National Institute for Environmental Studies (Environmental Observatory website: https://tenbou.nies.go.jp/download/).

The following items were extracted as patient profiles at the time of hospitalization and used as variables.

- Inpatient Facility/Infectious Disease Epidemic: Size of hospital beds, Type of facility (university, public, private), and period of infection spread.
- Patient attributes/ profile: gender, age, height, weight, BMI, smoking index, level of care required, and patient residence (city, ward, town, or village).
- Disease information: diabetes, hypertension, obesity, kidney disease, cardiovascular disease CVD, chronic obstructive pulmonary disease COPD comorbidity
- Air pollutants: Nitrogen dioxide NO2, Photochemical oxidant Ox, suspended particulate matter SPM, and fine particulate matter PM2.5 Concentrations for each substance in the month before the day of admission and cumulative concentrations for one year.

The prediction model was constructed using a machine learning random forest with a binary outcome of [death at discharge] or not. Python 3.9.7 was used as the program.

2. Results and Conclusions

Of the eligible cases, the valid data set consisted of 2,610 cases (126 deaths). The objective variable was death at discharge, which was split with the explanatory variables and down-sampled according to the number of cases for the objective variable. Next, a random forest model with 30 decision trees was created by splitting the data into 80% training data and 20% test data. The model had Accuracy of 0.80 (AUC:0.83) and the top features were age, Photochemical oxidant concentration one month prior to admission, and level of care needed. In addition, although the Accracy was lower for those aged 65 years and older (Accuracy:0.67, AUC:0.62), and the top feature values were cumulative SPM concentration one year prior to admission, cumulative NO2 concentration one year prior to admission, and cumulative PM2.5 concentration one year prior to admission. In a prognostic model for patients with Covid-19, we were able to show the influence of environmental factors as a feature from the patient profile at admission.

References

[1] Madjid M, Safavi-Naeini P, Solomon SD, Vardeny O. Potential Effects of Coronaviruses on the Cardiovascular System. JAMA Cardiology. 2020 March;5(7):831-840. doi:10.1001/jamacardio.

[2] Bozack A, Pierre S, et al. Long-Term Air Pollution Exposure and COVID-19 Mortality: A Patient-Level Analysis from New York City. Am J Respir Crit Care Med. 2022 March; 15:205(6):651-662.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230302

Prediction of Mental Health Support of Employee Perceiving by Using Machine Learning Methods

Hossein JAMALIRAD[a] and Mahdie JAJROUDI[a,1]
[a]*Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*

Abstract. Employees' mental health addresses concerns in the technology industry phenomenon. Machine Learning (ML) approaches show promise in predicting mental health problems and identifying related factors. This study used three machine learning models on OSMI 2019 dataset: MLP, SVM, and Decision Tree. Five features are extracted by permutation ML's method on the dataset. The results indicate that the models have been reasonably accurate. Moreover, they could effectively support predicting employee mental health comprehension in the technology industry.

Keywords. Machine Learning, Public Health, feature importance, prediction model

1. Introduction and Methods

In the new global economy, public mental health has become a central issue for interventions, resulting in a broad range of impacts and associated economic savings, even in the short term [1,2]. They can prevent any imposed cost from the individual up to firms, society, and governmental scope. Mental health assessments represent a framework and mechanism to fulfill reasonable responses for mentioned issue [3]. The lack of practical support for psychological problems can also affect an employee's productivity, absence, work commitment, and self-confidence [4]. Using machine learning techniques to predict mental health issues is considered a promising approach for early prevention of the issues mentioned [5]. The research is being conducted for using these methods to predict employees' perceiving of mental health support in the tech industry and ascertaining related factors. Although extensive research has been carried out on mental health [4,6-9], to our knowledge, no single study exists that focuses on the mental health understanding of employees.
We analyzed the Open Sourcing Mental Illness (OSMI) Mental Health in Tech Survey dataset[10]. Pre-processing techniques were adopted to handle both missing and inconsistent data. Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and Decision Tree were applied. To assess the validity of models used in this research, accuracy, sensitivity, and specificity are reported. Before the validity assessment, the

[1] Corresponding Author: Mahdie JAJROUDI, E-mail: jajroudi@gmail.com.

Permutation method was performed for feature selection. Needless to say, Scikit-learn 1.2 python library was used for all of these steps.

2. Result, Discussion and Conclusion

OSMI dataset expressing different views in 82 features and 352 respondents. The accuracy, sensitivity, and specificity of this model for MLP are (74.00%, 72.50%, and 75.42%); for decision-tree (83.62%, 78.72%, 88.57%) and finally, SVM (71.174%, 81.13%, 73.50%). Also, five final extracted features encompass.

This study used machine learning methods to predict mental health support for employees in the technology industry and determine related factors in the OSMI dataset. The results showed that the models were accurate enough to predict how much AI could help employers to know their employees, perceiving that they are receiving mental support in their technology industry. significant features included the effect of mental disorders on work, discussing mental health in the workplace, health raised in the interview, previous experiences, and, last, feedback from the employer. The accuracy of the obtained models can compete with studies in other fields, indicating a promising approach for mental health professionals and public health practitioners. Mainly, missing values and inconsistent data are limitations of this study.

The predicted results of the models examined in this research are of appropriate accuracy, indicating a promising approach for mental health professionals. Additionally, those who are active in the public health field figure out influential factors regarding the support value of employee understanding that they are receiving mental support in their technology industry.

References

[1] Alonso SG, et al. Data mining algorithms and techniques in mental health: a systematic review. Journal of medical systems. 2018; 42:1-15.

[2] Campion J. Public mental health: key challenges and opportunities. BJPsych international. 2018;15(3):51-4. doi.org/10.1192/bji.2017.11

[3] Brouwers E, et al. To disclose or not to disclose: a multi-stakeholder focus group study on mental health issues in the work environment. Journal of occupational rehabilitation. 2020; 30:84-92.

[4] Hogg B, , et al. Supporting employees with mental illness and reducing mental illness-related stigma in the workplace: an expert survey. European Archives of Psychiatry and Clinical Neuroscience. 2022.

[5] Gupta S, Saluja K, Goyal A, Vajpayee A, Tiwari V. Comparing the performance of machine learning algorithms using estimated accuracy. Measurement: Sensors. 2022;24: 100432.

[6] Alam T, Chen T, Bucholc M, Antoniou G. Investigating Mental Wellbeing in the Technology Workplace Using Machine Learning Techniques. Artificial Intelligence in Healthcare: Recent Applications and Developments: Springer; 2022. p. 165-77.

[7] Choi Y, Choi J. Job Satisfaction Prediction and Machine Learning Technique. 2022. doi.org/10.21203/rs.3.rs-1683972/v1

[8] Sujal B, et al., Mental Health Analysis of Employees using Machine Learning Techniques. 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS); 2022: IEEE. doi.org/10.1109/COMSNETS53615.2022.9668526

[9] Wei W. The Impact of Artificial Intelligence on the Mental Health of Manufacturing Workers: The Mediating Role of Overtime Work and the Work Environment. Frontiers in Public Health. 2022;10.

[10] Kaggle 2019 [Available from: https://www.kaggle.com/datasets/osmihelp/osmi-mental-health-in-tech-survey-2019.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230303

Improving Antibiotic Prescribing for Dentistry in France Using an Ontology

Jean-Alexis BUVAT[a,1], Francesco MONTI[a], Tariq NSAIS[a] and Bénédicte MELOT[a,b,c]
[a]*Institut des Sciences de la Santé Publique d'Aix-Marseille, ISSPAM, France.*
[b]*Université Sorbonne Paris Nord, LIMICS, Bobigny, France*
[c]*Qare, Paris, France*
ORCiD ID: Jean-Alexis BUVAT https://orcid.org/0000-0001-8111-2474,
Bénédicte MELOT https://orcid.org/0000-0001-7533-3924

Abstract. Antibiotic overprescribing in dentistry is a major concern that contributes to the emergence of antimicrobial resistance. It is due in part to the misuse of antibiotics by dentists but also by other practitioners who see patients in emergency for dental care. We used the Protégé software to create an ontology regarding the most common dental diseases and the most used antibiotics to treat them. It is an easy shareable knowledge base that could be used directly as decision support tool to improve the use of antibiotics in dental care.

Keywords. Antibiotics, Dentistry, Ontology, Emergency room decision support

1. Introduction

In France, misuse of antibiotics is seen in dentistry and in GP practice. About 60% of the prescriptions given by dentists worldwide are unnecessary or inappropriate [1]. Furthermore, many French people go to emergency services or a GP for urgent dental care [2] and 57.1% of GPs who see patients for dental emergencies prescribe antibiotics without waiting for a consultation with a dentist [3]. Indeed, antibiotics for urgent dental care are prescribed for 68% consultations on a Telemedicine platform and 33% at the ER (Melot B. et al, to be published). There are few existing ontologies regarding dental infections [4–6]. Our aim was to improve antibiotic prescribing for urgent dental care for French patients seen in primary care by the design of an ontology of dental infections.

2. Methods

To build the ontology on the topic of antibiotics in dentistry we used Protégé software [7] providing interoperability by supporting Web Ontology Language (OWL). The relevant concepts and relationships between antibiotics and diagnosis in dentistry were identified and set in the ontology using the latest French recommendations [8]. These included classes for different types of antibiotics, most common dental diseases and patient related knowledge and information (e.g. age, symptoms). Properties and relationships between the classes were defined. The ontology was continually reviewed

[1] Corresponding Author: Jean-Alexis BUVAT, ja.buvat@hotmail.fr.

as additional concepts and relationships were added. Subject matter experts were consulted and relevant literature was reviewed [9,10] to ensure the accuracy and completeness of the ontology.

3. Results and Discussion

We managed to create a knowledge base for dentistry in the form of an ontology regarding antibiotic, diseases and patient related information. It is based on 3 main classes: patient, dental disease and antibiotics. In each of those we implemented subclasses to be as exhaustive as possible (Figure 1).

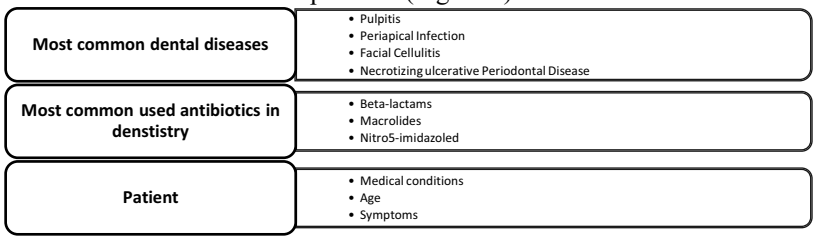


Figure 1. The three principal concepts and their subclasses in the ontology allowing the practitioner to identify the correct indication of antibiotics for dental care.

Despite current recommendations, patients are frequently prescribed antibiotics during acute teeth pain without clinical confirmation for infection. Our ontology is an exhaustive but simple tool allowing the mimic of the clinical reasoning for most common dental diseases. It organizes and structures up-to-date knowledge regarding antibiotic prescription in dentistry and improves the quality and safety of dental care. It does not exist yet in general practice in France and could be used for clinical decision-making algorithms or recommendation systems.

References

[1] Sukumar S, Martin FE, Hughes TE, Adler CJ. Think before you prescribe: how dentistry contributes to antibiotic resistance. Aust Dent J 2020;65:21–9. https://doi.org/10.1111/adj.12727.
[2] Etude Ifop / Dentaly.org. Les Français, l'accès aux soins dentaires et la prise en charge des prothèses dentaires 2018.
[3] Prise en charge des urgences bucco-dentaires en médecine générale : évaluation des pratiques en Ille-et-Vilaine n.d. https://syntheses.univ-rennes1.fr/search-theses/notice.html?id=rennes1-ori-wf-1-9365&printable=true (accessed March 5, 2023).
[4] Suzuki A, Takai-Igarashi T, Nakaya J, Tanaka H. Development of an Ontology for Periodontitis. J Biomed Semant 2015;6:30. https://doi.org/10.1186/s13326-015-0028-y.
[5] Bright TJ, Furuya EY, Kuperman GJ, Cimino JJ, Bakken S. Development and Evaluation of an Ontology for Guiding Appropriate Antibiotic Prescribing. J Biomed Inform 2012;45:120–8. https://doi.org/10.1016/j.jbi.2011.10.001.
[6] Shen Y, Yuan K, Chen D, Colloc J, Yang M, Li Y, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. Artif Intell Med 2018;86:20–32. https://doi.org/10.1016/j.artmed.2018.01.003.
[7] Musen MA. The Protégé Project: A Look Back and a Look Forward. AI Matters 2015;1:4–12. https://doi.org/10.1145/2757001.2757003.
[8] Antibiodent-2020_7.pdf n.d.
[9] Fehrencach MJ. Mosby's dental dictionary. vol. 4. St. Louis, Missouri : Elsevier; 2020.
[10] Rinzler CA. The encyclopedia of dental and oral health. New Yrok: Facts On File; 2011.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230304

Comparing Responses to COVID-19 Across Institutions: Conceptualization of an Emergency Response Maturity Model

Yalini SENATHIRAJAH[a,1], David KAUFMAN[b], Elizabeth BORYCKI[c], Andre KUSHNIRUK[c] and Kenrick CATO[b]
[a] *University of Pittsburgh School of Medicine, USA*
[b] *SUNY Downstate Health Sciences University, USA*
[c] *University of Victoria, Canada*
[d] *Columbia University, USA*
ORCiD ID: Yalini Senathirajah (https://orcid.org/0000-0001-8399-989X)

Abstract. The impact of Covid-19 on hospitals was profound, with many lower-resourced hospitals' information technology resources inadequate to efficiently meet the new needs. We interviewed 52 personnel at all levels in two New York City hospitals to understand their issues in emergency response. The large differences in IT resources show the need for a schema to classify hospital IT readiness for emergency response. Here we propose a set of concepts and model, inspired by the Health Information Management Systems Society (HIMSS) maturity model. The schema is designed to permit evaluation of hospital IT emergency readiness, permitting remediation of IT resources where necessary.

Keywords. Health IT, Emergency Preparedness, Maturity Model, COVID-19, Resilience

1. Introduction

Hospitals with fewer resources were often in neighborhoods with the most vulnerable patients and highest rates of infection, morbidity and mortality. In the US, 'safety net' hospitals (SNH) are a special designation of hospital which welcomes all patients' regardless of ability to pay. We studied a state-funded New York City hospital which was designated a COVID-19-only hospital, comparing it with a large networked hospital (LNH). The differences between the two institutions led us to propose a health IT emergency response maturity scale inspired by the HIMSS 7-level maturity model[1].

2. Methods

We conducted 52 sixty-minute zoom interviews with stakeholders including hospital leadership, clinical directors, IT staff, and others focused on decisions, tasks, and overall experience of organization resilience during COVID-19. The interviews were recorded,

[1] Corresponding Author: Yalini Senathirajah, E-mail: yalini@pitt.edu.

transcribed, and coded using MAXQDA™ qualitative analysis software and the Systems Engineering Initiative for Patient Safety (SEIPS)[2] schema. Data were mapped and analyzed for technologies used, automation and manual work, needs met and unmet.

3. Results

Experts identified the most significant emergent concepts and rankings describing practical needs, iterating over two cycles. In contrast to the LNH, the SNH had minimal electronic interconnection with other local hospitals. The staff overcame many technical and logistical lacks by sheer dedication, innovation, staff repurposing and clinical service change, which were critical to resilience. The proposed health IT emergency response maturity levels can be characterized in terms of three levels: 1) basic data integration, 2) information technology and 3) resilience capacity.

4. Discussion

The finding of major disparities in health IT resources and their critical role in controlling pandemics necessitates means of evaluating institutional readiness and health IT. We therefore propose an **'Emergency response maturity scale'** [HITERMS] for Health IT, akin to the Health Information Management Systems Society (HIMSS) technology maturity scale. This is a preliminary formulation based on empirical data and expert consultation. Our next step is to conduct a Delphi experiment with experts, IT personnel and management.

5. Conclusion

Health IT resource assessment is critical for future emergency preparedness, and the recent pandemic revealed many sharp disparities among institutions; these affect the entire ability of a society to respond adequately. A preparedness scale specifically addressing health IT disparities should be useful in conveying the extent of the problem and beginning to address these resource needs.

Acknowledgments

This work was funded by US AHRQ R01 HS028220 to Drs. Kaufman and Senathirajah.

References

[1] HIMSS Maturity Models. https://www.himss.org/what-we-do-solutions/digital-health-transformation/maturity-models. Accessed 3/8/22

[2] Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, Rivera-Rodriguez AJ. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics. 2013;56(11):1669-86.

[3] Senathirajah Y, Kaufman DK, Borycki, E, et al. The Impact of Regulatory Reporting Burden on a COVID-19-19 only Small Independent Hospital. Submitted to Health Affairs, March 2022

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230305

Opportunistic Screening for Osteoporosis Using CT Scans of the Knee: A Pilot Study

Mahmoud ELMAHDY[a] and Ronnie SEBRO[b,1]
[a] *Department of Radiology, Mayo Clinic, Jacksonville, Florida 32224, USA*
[b] *Department of Orthopedic Surgery, Mayo Clinic, Jacksonville, Florida, 32224, USA*
ORCiD ID: Mahmoud Elmahdy https://orcid.org/0000-0002-1363-0728.
Ronnie Sebro https://orcid.org/0000-0001-7232-4416.

Abstract. Knee CT scans are used for planning for total knee arthroplasties in patients who are often simultaneously at risk for frailty fractures due to low bone mineral density. We retrospectively identified 200 patients (85.5% female) with concurrent CT scans of the knee and Dual energy x-ray absorptiometry (DXA). The mean CT attenuation of the distal femur, proximal tibia and fibula, and patella, were calculated using volumetric 3-dimensional segmentation using 3D Slicer. Data were split randomly into training 80% and test 20% datasets. The optimal CT attenuation threshold for the proximal fibula was obtained in the training dataset and evaluated in the test dataset. A support vector machine (SVM) with radial basis function (RBF) using C-classification was trained and tuned using 5-fold cross-validation in the training dataset and then evaluated in the test dataset. The SVM had a higher area-under-the curve (AUC) of 0.937 and better performance to detect osteoporosis/osteopenia than the CT attenuation of the fibula (AUC of 0.717) (P=0.015). Opportunistic screening for osteoporosis/osteopenia could be accomplished using CT scans of the knee.

Keywords. Osteoporosis, computed tomography, Knee, Machine learning, Artificial Intelligence, Screening, Bone mineral density.

1. Introduction

We begin to lose bone mineral density in the third to fourth decades of life [1]. This is a pernicious process that eventually may lead to the development of low bone mineral density (BMD) [1]. Low BMD increases future fracture risk [2]. Hip fractures are associated with increased morbidity and approximately 33% mortality within 3 years after the hip fracture [1]. Dual energy X-ray absorptiometry (DXA) is the gold-standard for measuring BMD [1]. The World Health Organization (WHO) guidelines consider patients with lowest BMD T-score less than or equal to -2.5 as osteoporotic; patients with lowest BMD T-score between -2.5 and -1 as osteopenic; and patients with lowest BMD T-score greater than or equal to -1 as normal [1]. Although there are screening guidelines for low BMD using DXA, several eligible patients remain unscreened so there is a need for opportunistic screening using other methods [1]. Recent studies have shown that the computed tomography (CT) attenuation of trabecular bones, including the hand and wrist can be used to screen for low BMD [1]. We hypothesize that knee CT scans performed

[1] Corresponding Author: Ronnie Sebro MD, PhD, E-mail: rsebro@gmail.com.

as part of routine clinical practice including knee arthroplasty [2] could be used to screen for low BMD, and we hypothesize that machine learning techniques [1] [3] utilizing the CT attenuation from several bones would be better than using the CT attenuation at a single bone.

2. Materials and Methods

This is a retrospective study of patients evaluated between 01/01/2013 and 12/31/2021 at a tertiary care academic center who had knee CT and DXA scans performed within 12 months of each other. Patients were excluded if they already had knee arthroplasties or fractures. 3D Slicer was used to segment the epiphysis/metaphysis of the distal femur, proximal tibia, and proximal fibula; and the patella to obtain the CT attenuation of these bones. The dataset was randomly split into training/validation (80%) and test (20%). First, the optimal threshold CT attenuation of the fibula to predict whether a patient had osteoporosis or osteopenia was identified in the training dataset and evaluated in the test dataset. Next, we used a support vector machine (SVM) with radial basis function and C-classifier to predict whether a patient had osteoporosis or osteopenia because this model was effective in a prior paper [1]. Tuning was done with 5-fold cross-validation, using epsilon ranging from 0 to 1 in increments of 0.1, and cost of 1 to 8 in unit increments. The performance of the optimal tuned SVM was evaluated in the test dataset. DeLong's test was used to compare the area under the curve (AUC) of the SVM to the CT attenuation of the fibula in the test dataset. Statistics were two-sided with alpha set at 0.05. Statistics were performed using Rv4.04.

3. Results and Conclusions

A total of 200 patients were identified, 171 (85.5%) women and mean age (standard deviation) of 67 (8.0) years. A threshold CT attenuation of the fibula of 45.8 Hounsfield Units (HU) had a sensitivity of 0.733, specificity of 0.700, and AUC of 0.717, while the SVM had a sensitivity of 0.833, specificity of 1.00, and AUC of 0.937 to predict osteoporosis or osteopenia in the test dataset. The SVM classifier had superior performance to the single CT attenuation measurement at the fibula (P=0.015).
Opportunistic screening for osteoporosis and osteopenia can be performed from CT scans of the knee. Future research is required to evaluate how fracture risk can be estimated from CT scans of the knee.

References

[1] Sebro R, De la Garza-Ramos C. Machine Learning for Opportunistic Screening for Osteoporosis from CT Scans of the Wrist and Forearm. Diagnostics (Basel). 2022 Mar 11;12(3):691. doi: 10.3390/diagnostics12030691.

[2] Bernatz JT, Krueger DC, Squire MW, Illgen RL 2nd, Binkley NC, Anderson PA. Unrecognized Osteoporosis Is Common in Patients With a Well-Functioning Total Knee Arthroplasty. J Arthroplasty. 2019 Oct;34(10):2347-2350. doi: 10.1016/j.arth.2019.05.041. Epub 2019 May 30. PMID: 31227302.

[3] Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ. Current Applications and Future Impact of Machine Learning in Radiology. Radiology. 2018 Aug;288(2):318-328.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230306

Opportunistic Screening for Osteoporosis Using Hand Radiographs: A Preliminary Study

Farid Ghareh MOHAMMADI[a,b] and Ronnie SEBRO[a,b,1]
[a] *Department of Radiology, Mayo Clinic, Jacksonville, Fl, 32224, USA*
[b] *Center for Augmented Intelligence, Mayo Clinic, Jacksonville, Fl, 32224, USA*

Abstract. Patients with low bone mineral density (BMD) are at risk for fractures however are often undiagnosed. Therefore, there is a need to opportunistically screen for low BMD in patients who present for other studies. This is a retrospective study of 812 patients aged 50 years or older who had dual-energy X-ray absorptiometry (DXA) and radiographs of the hands within 12 months of each other. This dataset was randomly split into training/validation (n=533) and test (n=136) datasets. A deep learning (DL) framework was used to predict osteoporosis/osteopenia. Correlations between the textural analysis of the bones and DXA measurements were obtained. We found that the DL model had an accuracy of 82.00%, sensitivity of 87.03%, specificity of 61.00% and an area under the curve (AUC) of 74.00% to detect osteoporosis/osteopenia. Our findings show that radiographs of the hand can be used to screen for osteoporosis/osteopenia and identify patients who should get formal DXA evaluation.

Keywords. Deep Learning, Radiograph, DXA, Bone Mineral Density, Hand

1. Introduction

Bone mineral density (BMD) loss begins in the fourth decade of life and increases significantly after menopause for women [1] [2]. Low BMD increases the risk of fractures [1] [2]. Dual-energy X-ray absorptiometry (DXA) is the gold standard test for the evaluation of BMD [3]. Most patients who are at risk for low BMD are often not screened using DXA, so there is a need for opportunistic screening for low BMD [1] [2]. A recent study showed that radiographs of the hip and lumbar spine could be used to predict BMD [4]. We hypothesized that hand radiographs could also be used to predict BMD, so the goal of the study was to use hand radiographs for opportunistic screening for low BMD to identify patients who should have a formal evaluation with DXA.

2. Methods

We conducted a retrospective cohort study of patients at Mayo Clinic aged >=50 years with DXA and hand radiographs taken within 12 months of each other between 01/01/2010 and 12/31/2021.

[1] Corresponding Author: Ronnie Sebro, E-mail: Sebro.ronnie@mayo.edu.

2.1. Image acquisition and image pre-processing

Radiographs were obtained using Siemens YSIO (Siemens Healthineers, Erlangen, Germany) radiographic machines. Posterior-anterior, oblique, and lateral radiographs were obtained. The patients were randomly split into training/validation (80%) and test (20%) datasets. Images were converted into Joint Photographic Experts Group (JPEGs /JPGs) format, resized to 256 x 256, followed by image intensity normalization.

2.2. CNN architecture

We used Python 3.9.15 to create a deep learning model (four convolutional neural network (CNN) layers along with three layers of fully connected neural network (FCNN)) using 2388 hand radiographs from 533 patients to predict whether a patient had osteopenia/osteoporosis. The cross-entropy loss function was minimized. The model was run over 50 epochs using Adam optimization. Five-fold cross-validation was used to tune the model and the optimal tuned model was evaluated in the test dataset. Sensitivity, specificity, accuracy, area under the curve (AUC), positive predictive value (PPV), and negative predictive value (NPV) were calculated.

3. Results

The study comprised of 669 patients (84% women) with mean age of 65.48 (50-91) and stand deviation of 8.90. 17.60% of patients were osteoporotic; 59.00% were osteopenic; and 23.40% had normal BMD. When predicting osteopenia/osteoporosis, the optimal model had sensitivity of 93.45%, specificity of 97.05%, accuracy of 94.37%, AUC of 95.25%, PPV of 98.91%, and NPV of 83.54% in the training dataset and sensitivity of 87.03%, specificity of 61.00%, accuracy of 82.00%, AUC of 74.00%, PPV of 89.52%, and NPV of 55.00% in the test dataset.

4. Conclusion

Opportunistic screening for low BMD can be done using deep learning models evaluating hand radiographs. Hand radiographs can be used to identify patients who should go on to get screening DXA studies. Further research is needed to assess whether deep learning models could also predict future fracture risk from radiographs.

5. References

[1] Warming L, Hassager C, Christiansen C. Changes in bone mineral density with age in men and women: a longitudinal study. Osteoporosis International. 2002;13:105-12.
[2] Sebro R, De la Garza-Ramos C. Machine learning for opportunistic screening for osteoporosis from CT scans of the wrist and forearm. Diagnostics. 2022;12(3):691.
[3] Bachrach L. Dual energy X-ray absorptiometry (DEXA) measurements of bone density and body composition: promise and pitfalls. Journal of pediatric endocrin & metabol: JPEM. 2000;13:983-8.
[4] Hsieh C-I, Zheng K, et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. Nature communications. 2021;12(1):5472.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230307

Linkage Health and Environmental Data: A Case Study on Asthma Prevalence in Children and Adolescents in Slovenia

Tanja REJC[a,1], Andreja KUKEC[a,b], Tanja CARLI[b] and Ajda MLAKAR[a]
[a] *Faculty of medicine, Department of Public Health, University of Ljubljana, Slovenia*
[b] *National Institute of Public Health, Slovenia*
ORCiD ID: Andreja Kukec https://orcid.org/0000-0002-5973-0345

Abstract. For the last 10 years there is no data on the prevalence of asthma in Slovenian children. To ensure accurate and high-quality data we will conduct a cross-sectional survey Health Interview (HIS) and Health Examination Survey (HES) design type. Therefore, we first prepared the study protocol. To get the data for the HIS part of the study we developed a new questionnaire. The exposure to outdoor air quality will be evaluated from the National Air Quality network data. In Slovenia the problems with health data should be addressed with the common unified system at the national level.

Keywords. Protocol, prevalence study, asthma in children, health data, environmental data

1. Introduction

Asthma is the most common chronic disease, and cause of hospitalization of children. Researchers have linked air pollutants to a number of respiratory diseases [1]. For the last 10 years there is no data on the prevalence of asthma in Slovenian children. Environmental history questionnaire which could effectively assess the impact of environmental risk factors on the occurrence or worsening of asthma in children, has not yet been developed. The aim of the study is to design the methodology to get the best data possible to assess the national asthma prevalence in children and adolescents and its associated environmental risk factors.

2. Methods

Prior conducting the study the protocol was developed. The observed outcome will be asthma prevalence estimated using HIS and HES approach. To achieve the HIS part of the study a new questionnaire was developed considering 3 previously used questionnaires, aiming the assessment of respiratory diseases in relation to the built environment [2-4]. Observed population group will be 6 to 7 years old children, and 12 to 13 years old adolescents in 75 randomly selected primary schools in Slovenia.

[1] Corresponding Author: Tanja Rejc, E-mail: Tanja.rejc@mf.uni-lj.si.

3. Results

Based on the study protocol a set of questions considering our aim of research was selected from each of the three mentioned questionnaires. Together with the clinic pediatric specialists we added and reformulated questions, answers and certain terms to target more accurately the observed outcome and environmental risk factors. Validation process will be performed prior national cross-sectional study. The exposure to outdoor air quality will be additionally evaluated from the National Air Quality network data, considering the child's residential address. The cases with parent-reported current asthma, and the cases where the child's symptoms indicate possible presence of asthma, will be in the next step examined and confirmed by a pediatrician reviewing the child's medical records. The obtained data will be furthermore organized into a single database for statistical analyses.

4. Discussion

Kukec and colleagues report a significant differences and discrepancies in recording the health data between the health centres, and inconsistencies in using the International Classification of Diseases [3]. As we expected the same problems in our study, we formulated different methodological approach to collect the health data. The combination of HIS and HES type of study ensure us the verified and good quality data [5].

5. Conclusion

To get the quality data on asthma prevalence and asthma exacerbation in relation to environmental factors the study HIS and HES type is necessary. In Slovenia, there is a need to unify the methodology regarding the health data recording. The creation of a common database in which already obtained data on environmental pollution and health could be linked together would be of extreme importance in enduring the quality of public health research.

References

[1] Khreis H, Kelly C, Tate J, Parslow R, Lucas K, Nieuwenhuijsen M. Exposure to traffic-related air pollution and risk of development of childhood asthma: A systematic review and meta-analysis. Environ Int. 2017; 100:1-31.

[2] Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. Lancet. 1998;25;351(9111):1225-32.

[3] Kukec A, Zaletel-Kragelj L, Farkaš Lainščak J, et al. Health geography in case of Zasavje: Linking of atmospheric air pollution and respiratory diseases data. Acta Geographica Slovenica, 2014 54(2), 345–362. https://doi.org/10.3986/AGS54208

[4] Szabados M, Csákó Z, Kotlík B, et al. Indoor air quality and the associated health risk in primary school buildings in Central Europe - The InAirQ study. Indoor Air. 2021;31(4):989-1003.

[5] Scheidt-Nave C, Kamtsiuris P, Gößwald A, et al. German health interview and examination survey for adults (DEGS) - design, objectives and implementation of the first data collection wave. BMC Public Health. 2012(1); 12:730.

Section 9

Telehealth, Sensors, Signals and Imaging Informatics

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230309

Towards Automated COVID-19 Presence and Severity Classification

Dominik MUELLER[a,b], Silvan MERTES[a], Niklas SCHROETER[a], Fabio HELLMANN[a], Miriam ELIA[a,1], Bernhard BAUER[a], Wolfgang REIF[a], Elisabeth ANDRÉ[a] and Frank KRAMER[a]
[a]*Faculty of Applied Computer Science, University of Augsburg, Germany*
[b]*Institute for Digital Medicine, University Hospital Augsburg, Germany*

Abstract. COVID-19 presence classification and severity prediction via (3D) thorax computed tomography scans have become important tasks in recent times. Especially for capacity planning of intensive care units, predicting the future severity of a COVID-19 patient is crucial. The presented approach follows state-of-theart techniques to aid medical professionals in these situations. It comprises an ensemble learning strategy via 5-fold cross-validation that includes transfer learning and combines pre-trained 3D-versions of ResNet34 and DenseNet121 for COVID19 classification and severity prediction respectively. Further, domain-specific preprocessing was applied to optimize model performance. In addition, medical information like the infection-lung-ratio, patient age, and sex were included. The presented model achieves an AUC of 79.0% to predict COVID-19 severity, and 83.7% AUC to classify the presence of an infection, which is comparable with other currently popular methods. This approach is implemented using the AUCMEDI framework and relies on well-known network architectures to ensure robustness and reproducibility.

Keywords. COVID-19, Deep Learning, Severity, Classification, Infection-LungRatio, Ensemble Learning, AUCMEDI

1. Introduction

In response to the rapid spread of the SARS-CoV-2 virus at the beginning of the year 2020, many scientists quickly reacted and developed various approaches based on deep learning to contribute to the efforts against COVID-19. Especially the severity assessment of patients is essential for treatment decisions and disease course monitoring. However, the course of the SARS-CoV-2 virus pandemic showed that one of the most critical factors for COVID-19 treatment is the capacity of intensive care units (ICU) at hospitals [1,2]. Through the rapid but inconsistent development of infection severity, predicting the future severity of a patient within a month (prognosis) for capacity planning of ICUs is challenging but crucial [1,2]. This paper presents an approach to assess COVID-19 severity through the classification of 3D thorax computed tomography (CT) scans. It relies on the in-house AUCMEDI framework, which provides easy access to pretrained models, ready for medical transfer learning. This approach enables swift

[1] Corresponding Author: Miriam Elia, E-mail: miriam.elia@informatik.uni-augsburg.de.

adoption since the heavy lifting is moved to the framework. Our method achieves promising results, and we encourage other researchers to build on the presented baseline.

2. Method

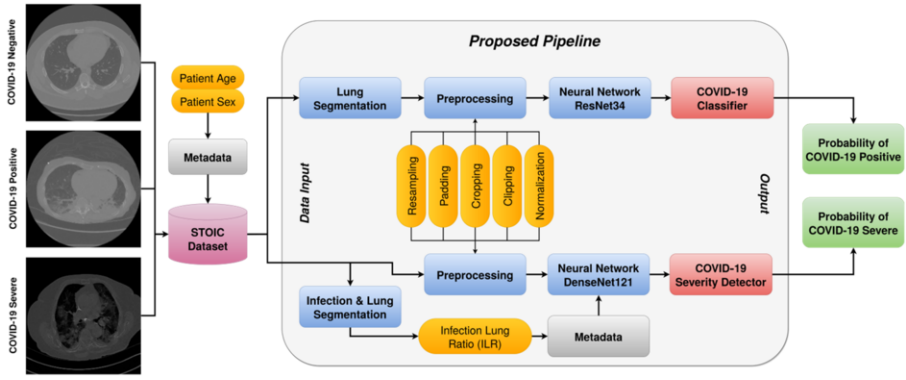


Figure 1. Flowchart diagram of the proposed pipeline.

To set up a modern and effective medical image classification pipeline (Figure 1), the in-house framework AUCMEDI was utilized [3], a software package that offers a library as a high-level API for the standardized construction of modern medical image classification pipelines. As illustrated in Figure 1, our model takes thorax CT scans as input, in addition to the patient's age, sex, and *Infection-Lung Ratio* (ILR), which is computed a priori. The outputs of the model are estimated probabilities for COVID-19 presence and a severe outcome within a month. An ensemble approach was applied to assess the presence and severity of COVID-19 based on CT scans of the thorax, in which independent models were used for the individual prediction of presence with our 'COVID-19 Classifier' and severity with our 'COVID-19 Severity Detector'. For model training and internal performance evaluation, the dataset from the STOIC (Study of Thoracic CT in COVID-19) project was used [4].

2.1. Pre-processing and Image Augmentation

The following image augmentation methods were applied for the training process: rotation, flipping, scaling, gamma modification, and elastic deformations. In contrast to the COVID-19 Severity Detector, a segmentation of the lung is applied for the COVID19 Classifier workflow before pre-processing. Pixels outside of the segmented lung region are excluded and the volume is cropped to a minimal shape around the lung. For the pre-processing of both models, the CT volume is resampled to a voxel spacing of 1.48x1.48x2.10 mm and clipped to the range −1024 Hounsfield Units (HU) to +100 HU to exclude irrelevant tissue types as well as to reduce complexity. Subsequently, the pixel intensities of the volume were standardized to a grayscale range. Samples that might exceed the accepted input image size of 148x224x224 pixels are center-cropped or

padded if undersized to match the required image size. For training, random cropping was applied instead of center-cropping. As the last pre-processing step, another normalization is applied via the Z-Score normalization approach based on the mean and standard deviation computed on the ImageNet dataset [5]. The severity of lung diseases is commonly quantified in radiology by the percentage of lung area which is affected. This ratio between affected and healthy lung tissue provides intuitive insights into the current state and severity of the disease. For pneumonia as well as COVID-19, a similar approach can be applied. The ILR describes the ratio between infected and healthy tissue in the lung measured in pixels: $ILR = |Infection| / (|Infection| + |Lung|)$. Several studies [6-8] proved that the ILR assessed on CT scans has a high correlation to severity measured by labor biomarkers as well as the survival rate. Thus, the computed ILR was integrated as metadata in the COVID-19 Severity Detector, utilizing the pipeline by [9].

2.2. Neural Network Models and Ensemble Learning Strategy

For the COVID-19 Classifier, a 3D version of the ResNet34 [10] architecture is used, whereas the COVID-19 Severity Detector is based on a 3D version of the DenseNet121 [11] architecture. The COVID-19 Severity Detector classification head was modified to additionally take metadata into account. The metadata consists of three parts: Patient age, sex, and the ILR of each sample. Regarding the training process of both models, transfer learning was applied to the classification head, and a fine-tuning strategy on all layers. The transfer learning was conducted for 10 epochs, using the Adam optimizer with an initial learning rate of 1×10^{-4} and a batch size of 4 for the DenseNet, and 8 for the ResNet. The fine-tuning run for a maximum of 240 epochs, using a dynamic learning rate starting from 1×10^{-5} to a maximum decrease of 1×10^{-7} (decreasing factor of 0.1 after 8 epochs without improvement on the monitored validation loss). Furthermore, an early stopping technique was utilized, stopping after 36 epochs without improvement. As a loss function, the sum of the F1-score and the weighted Focal loss from Lin et al. [12] was utilized. For inference, the model with the best validation loss is used. The COVID-19 Classifier predicts presence and severity, resulting in the following three classes: Negative, Positive, and Severe (subset of Positive).

A 5-fold cross-validation was applied on both models as a Bagging approach for ensemble learning. For the performance evaluation, the combined cross-validation folds for training were split into a training (70%) and validation subset (10%), whereas the remaining cross-validation fold (20%) was utilized for testing.

The prediction of the final COVID-19 Severity Detector pipeline comprises the mean-averaged sum of all five predictions from the five models of the cross-validation. This approach not only allows for more efficient usage of the available data but also increases the reliability of the prediction. However, for the COVID-19 Classifier, only the model with the best-monitored validation loss was used for predicting the COVID-19 Positive class in the final pipeline. Internal experiments revealed that this allowed achieving the best testing performance for COVID-19 presence prediction compared to utilizing a similar pooling-based prediction strategy like the COVID-19 Severity Detector.

3. Results

By utilizing cross-validation for evaluation, we were able to compute performance metrics on all samples (Table 1). For the COVID-19 Classifier, a multi-class evaluation was performed in which the class with the highest predicted probability was used as the outcome of the evaluation. As metrics, the Accuracy, F1-Score, and AUC were computed.

Classifier			
Classes	Acc.	F1	AUC
Negative	0.787	0.782	0.837
Positive	0.701	0.629	0.766
Severe	0.782	0.424	0.788
Classifier and Severity Detector			
Method	Probability		AUC
Classifier	infection		0.837
Detector	severe outcome		0.790

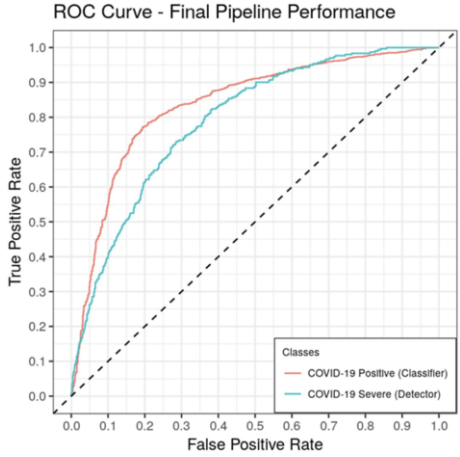


Table 1. Computed performance results of the Figure 2. Results of the performance evaluation for COVID-19 Classifier and Severity Detector. COVID-19 infection and severe outcome prediction.

For the final performance assessment, the predicted probabilities for the COVID-19 infection and severe outcomes were evaluated by the AUC metric. The COVID-19 infection prediction achieved an AUC of 83.7%, whereas the severe COVID-19 outcome prediction had an AUC of 79.0% (Figure 2). In contrast, the COVID-19 Classifier obtained an AUC of 78.8% for severity outcome prediction.

4. Discussion

The proposed pipeline demonstrated robust classification performance for predicting COVID-19 infection and severe outcomes within a month. This is why a separate model for severity prediction with the COVID-19 Severity Detector was implemented instead of utilizing the inferior COVID-19 Severe class prediction of the COVID-19 Classifier. Overall, accurately predicting the severe outcome within a month is challenging. Analyzing the prediction results of the COVID-19 Classifier revealed that especially the differentiation between COVID-19 Positive and COVID-19 Severe patients is a hard and complex task.

5. Conclusions

In this study, a powerful pipeline for COVID-19 infection and severe outcome prediction was implemented, utilizing the AUCMEDI framework combined with integrated

ensemble, transfer, and deep learning techniques. The pipeline consisted of a multi-model workflow for individual prediction of the outcome variables to ensure utilization of the best-suited model design. By utilizing a standardized framework, our approach is easily reproducible which guarantees a fast adoption for other projects. In future work, additional validation and integration in a clinical study are needed to identify actual medical gain in a clinical workflow.

References

[1] Silvani P, Colombo S, Scandroglio AM, Dell'Acqua A, Fominskiy E, Monti G, et al. Fast reshaping of intensive care unit facilities in a large metropolitan hospital in Milan, Italy: facing the COVID-19 pandemic emergency. Critical Care and Resuscitation. 2020 4;22(2):91-4.

[2] Goh KJ, Wong J, Tien JCC, Ng SY, Duu Wen S, Phua GC, et al. Preparing your intensive care unit for the COVID-19 pandemic: Practical considerations and strategies. Critical Care. 2020 5;24(1):215.

[3] Mayer S, Muller D, Kramer F. Standardized Medical Image Classification across¨ Medical Disciplines. arXiv; 2022. Available from: https://arxiv.org/abs/ 2210.11091.

[4] Revel MP, Boussouar S, de Margerie-Mellon C, Saab I, Lapotre T, Mompoint D, et al. Study of Thoracic CT in COVID-19: The STOIC Project. Radiology. 2021 10;301(1):E361-70.

[5] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision. 2015 12;115(3):211-52.

[6] Wang X, Che Q, Ji X, Meng X, Zhang L, Jia R, et al. Correlation between lung infection severity and clinical laboratory indicators in patients with COVID-19: a cross-sectional study based on machine learning. BMC Infectious Diseases. 2021 12;21(1):192.

[7] Xiong Y, Sun D, Liu Y, Fan Y, Zhao L, Li X, et al. Clinical and High-Resolution CT Features of the COVID-19 Infection: Comparison of the Initial and Follow-up Changes. Investigative Radiology. 2020 6;55(6):332-9.

[8] Raoufi M, Ahmad Safavi Naini SA, Azizan Z, Jafar Zade F, Shojaeian F, Boroujeni MG, et al. Correlation between Chest Computed Tomography Scan Findings and

[9] Mortality of COVID-19 Cases; a Cross sectional Study. Archives of Academic Emergency Medicine. 2020 5;8(1):e57.

[10] Muller D, Soto-Rey I, Kramer F. Robust chest CT image segmentation of COVID-¨ 19 lung infection based on limited data. Informatics in Medicine Unlocked. 2021 1;25:100681.

[11] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). vol. 2016-Decem. IEEE Computer Society; 2016. p. 770-8.

[12] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700-8.

[13] Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object De-´ tection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017 8;42(2):318-27.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230310

Designing a Persuasive E-Coaching Application for Informal Caregivers

Shweta PREMANANDAN[a], Awais AHMAD[b,1], Åsa CAJANDER[a] and Pär Ågerfalk[a]
[a] *Department of Informatics and Media, Uppsala University, Sweden*
[b] *Department of Information Technology, Uppsala University, Sweden,*

Abstract. Being an informal caregiver is not easy, and might cause physical and psychosocial burden, especially in the long run. However, the formal health care system has little support for informal caregivers who experience abandonment and lack information. Mobile health can potentially be an efficient and cost-effective way of supporting informal caregivers. However, research has shown that mHealth systems often have problems with usability, and people do not use the systems for more than a short period. Therefore, this paper explores the design of an mHealth app using Persuasive Design, an established design framework. This paper presents the design of the first version of the e-coaching application using the persuasive design framework and unmet needs of informal caregivers from the literature. This prototype version will be updated based on interview data from informal caregivers in Sweden.

Keywords. eHealth, Informal caregiving, persuasive design, e-coaching, mHealth

1. Introduction

Informal caregivers (hereafter referred to as caregivers) provide in-home care to relatives who are sick or injured [1]. This informal caregiving is an essential social responsibility many people take on voluntarily. However, caregivers can suffer from adverse outcomes like stress and anxiety, and the many risks of being a caregiver have been extensively studied [2]. The formal healthcare structure is stretched in terms of resources such as money and human resources and cannot provide sustainable long-term care needs of patients. There is indeed a trend in health care to provide more care in the home due to an ageing population, and home monitoring systems, smart homes, and connected health are areas that have emerged as a result. In parallel, studies show people have difficulty using mHealth systems, especially among older adults [3,4].

One promising avenue in the quest for successful eHealth design is the area of design methodologies for eHealth, such as the Persuasive System Design Model (PSDM). When using a structured way of designing an eHealth application, it is more likely that it has a functional and efficient design [5]. PSDM has shown promising results related to, for example, physical activity [6]. However, it has rarely been used in the area of informal caregiving. In this paper, we have used PSDM to design the first version of an e-coaching application for caregivers called AnhörigCare by following the persuasive design principles, and considering caregivers' unmet needs from the literature. This paper

[1] Corresponding Author: Awais AHMAD, E-mail: awais.ahmad@it.uu.se.

contributes by illustrating a systematic approach to designing a persuasive e-coaching application for caregivers using PSDM. This study uses insights from previous research on informal caregiving contexts and their needs for an e-coaching application (AnhörigCare). Previous research emphasises that participants may have difficulties conceptualizing a mobile app for caregiving, hence, we design a prototype using literature that will further be updated through user data [7].

AnhörigCare is an e-coaching application that aims to support caregivers in Sweden in their caregiving activities, along with assisting them in self-care. It will provide a place where caregivers can have access to practical information and access formal services regarding caregiving, including companion service, booking sessions with a therapist, and more. In the following section, we provide an overview of the persuasive designing process and how it is used to design the first version of AnhörigCare in this study.

PSDM provides a systematic approach to designing engaging and applicable interventions with greater sustainability [8]. It proposes three main design steps: analysing the main aspects of a persuasive system, understanding the persuasion context, and designing system features [5]. PSDM then provides 28 design principles grouped into four dimensions. 1) primary task support that helps users perform their target behaviours, 2) dialogue support that uses design principles that motivate users through feedback and interaction with the app, 3) credibility support uses techniques that make applications look and feel trustworthy to users, and finally, 4) social support, which uses techniques that leverage social influence [5]. In this paper, we illustrate the process of designing the first version of AnhörigCare (see Figure 1).

Figure 1. PSD process for AnhörigCare

2. Method

Here we applied the persuasive design process to AnhörigCare as illustrated in figure 1 by beginning to review the literature to understand the persuasion context, including the purpose that the system will serve, the context in which the system will be used along with the technological context, and the persuasion strategy to be used.

2.1. Review of Literature

A scoping review is being conducted using the Arksey and O'Malley framework [9]. From an initial screening of the papers (56 in total), we identified broad categories of caregiver needs which are used in this study. The extant literature points to access to information regarding caregiving, access to formal services to assist caregivers, feeling of community [10], words of acknowledgment and encouragement, self-care [11], and informal peer support [12] as major needs of caregivers. The descriptions of their needs were compared with the persuasive design principles from PSDM. Based on this match, a design principle was chosen to meet their needs.

2.2. Analysis of Persuasion Context

This application aims to support caregivers to continue providing care while encouraging them to engage in self-care. AnhörigCare will be designed to autogenously enable caregivers to change their behaviour and attitude without healthcare professionals' active participation. The application will provide one-stop access to practical information and formal support available to caregivers in the applications. Caregivers could access information from experts in a format (text, video clips, etc.) and in a language of their choice regarding caregiving. They would also be able to access additional formal support services provided by the municipalities for caregivers. Additionally, caregivers would be able to set self-care goals and be motivated by the application to follow through. AnhörigCare will be designed as a mobile application. Both direct and indirect messaging routes will be used. Practical information about caregiving activities and access to formal services is provided using a direct route, while goal-setting and self-monitoring of activities for self-care are provided using an indirect route.

3. Results

Here we present the persuasion features of the first version of AnhörigCare which is designed with the help of identified needs in the literature using an interactive prototyping tool [13]. These persuasion features and their implementation are summarized in Table 1.

Table 1. Design of System Qualities

Category	Feature	Implementation
Primary task Support	Tailoring	The application is designed to provide tailored data based on their location, the condition they provide care for, and their relationship to the care recipient (refer Fig. 2, needs for information and formal services).
	Self-monitoring	Provides a provision to connect other health apps and health devices and monitor their progress (refer to Fig. 4, need for self-care).
	Goal-setting	Allows caregivers to set their weekly or daily goals and follow through (refer to Fig. 3, need for self-care).
Dialogue Support	Liking	Use of words and phrases that caregivers can identify with. For e.g., 'AnhörigCare' means caring for 'anhöriga', the Swedish word for relatives.
	Praise	Use motivational quotes and encouraging textual feedback after the completion of a day of caregiving
	Reminder	Caregivers can set reminders for time-sensitive care activities like giving medicines to their care recipient, cleaning etc. Also allows them to set reminders about their self-care (refer Fig. 5).
Credibility Support	Expertise	Provides contact information of caregiver support organizations like Anhörigas Riksförbund counsellors and links to reach their homepage (refer Fig. 6).
	Trustworthiness	Provides information about and contact information for the team involved in developing the application.
Social Support	Social learning	Provide a discussion forum where caregivers can discuss their experiences or challenges (refer Fig. 7, informal peer support, feeling of community).

Cooperation	Via a discussion forum, caregivers can feel that they are part of a community with a common interest (refer Fig. 7, informal peer support, feeling of community).

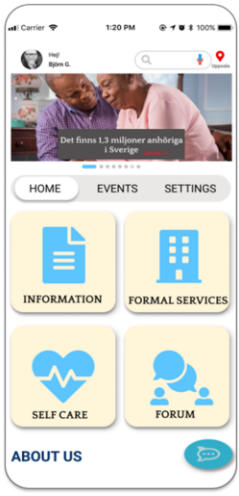


Figure 2. Home Page

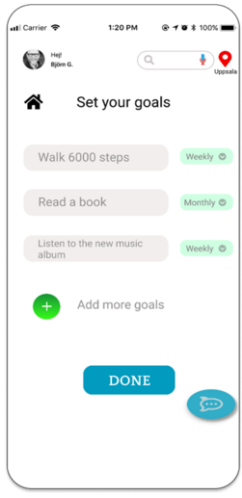


Figure 3. Goal-setting

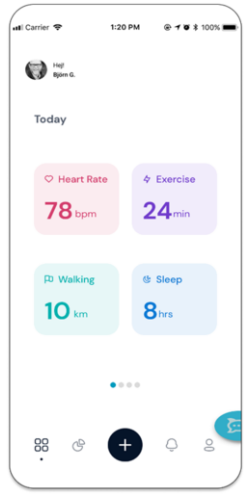


Figure 4. Self-monitoring

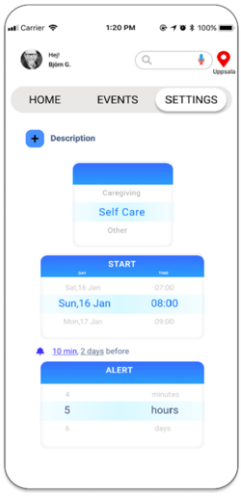


Figure 5. Reminders

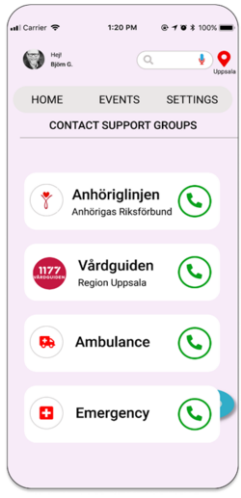


Figure 6. Contact information

Figure 7. Online Forums

4. Discussion and Conclusion

This study uses insights from previous research on informal caregiving contexts and their needs for an e-coaching application (AnhörigCare). Previous research gave us an understanding of caregivers' needs of caregiving to patients such as tailored and trustable information, access to formal services, caregivers' self-care and well-being,

acknowledgment, social interaction, and engagement [10–12]. These needs are addressed using PSDM making AnhörigCare more usable and engaging. Interestingly, we found the most suitable persuasive design principles to address the caregivers' needs. PSDM helps in designing effective and usable persuasive e-health applications for caregivers [5] through Primary Task Support (e.g., tailored practical information, monitoring health, and setting goals), Dialogue Support (acknowledgments and praise, and setting reminders), Credibility Support (quick access to caregiver support organizations and information on design team), and Social Support (social learning, and cooperation through online forums).

Based on PSDM and the unmet needs of caregivers from previous literature, we designed an initial prototype of AnhörigCare. This prototype might be useful for Information Systems researchers and designers as a first step in creating meaningful applications with a focus on elements that nudge users towards effective use of applications. A limitation of this paper is the lack of empirical evidence from caregivers in Sweden. As the involvement of actual users in design and development is of utmost importance for better usability and usefulness [1], our next step is to obtain empirical evidence by interviewing caregivers in Sweden. The next version of the prototype will be evaluated with caregivers which is a limitation in the current study. It will be interesting to understand how our findings from this study compare to the empirical evidence from the Swedish context.

References

[1] Ahmad A, Cajander Å, Johansson B, Tiblom Ehrsson Y, Langegård U. Designing for Human Well-Being: A Case Study with Informal Caregivers of Individuals with Cancer. Chall Trust AI Added-Value Health. 2022;214–8.
[2] Beach SR, Schulz R, Williamson GM, Miller LS, Weiner MF, Lance CE. Risk Factors for Potentially Harmful Informal Caregiver Behavior. J Am Geriatr Soc. 2005;53(2):255–61.
[3] Hallewell Haslwanter JD, Fitzpatrick G. The development of assistive systems to support older people: issues that affect success in practice. Technologies. 2017;6(1):2.
[4] Landgren S, Cajander Å. Non-use of Digital Health Consultations Among Swedish Elderly Living in the Countryside. Front Public Health [Internet]. 2021 [cited 2023 Feb 23];9. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2021.588583
[5] Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. Commun Assoc Inf Syst. 2009;24(1):485–500.
[6] Aldenaini N, Alqahtani F, Orji R, Sampalli S. Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. Front Artif Intell. 2020;3:7.
[7] Köhle N, Drossaert CH, Oosterik S, Schreurs KM, Hagedoorn M, van Uden-Kraan CF, et al. Needs and Preferences of Partners of Cancer Patients Regarding a Web-Based Psychological Intervention: A Qualitative Study. JMIR Cancer. 2015 Dec 29;1(2):e13.
[8] Oinas-Kukkonen H, Harjumaa M. A Systematic Framework for Designing and Evaluating Persuasive Systems. In 2008. p. 300.
[9] Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. Int J Soc Res Methodol Theory Pract. 2005;8(1):19–32.
[10] Plöthner M, Schmidt K, de Jong L, Zeidler J, Damm K. Needs and preferences of informal caregivers regarding outpatient care for the elderly: a systematic literature review. BMC Geriatr. 2019 Mar 13;19(1):82.
[11] Angelo J, Egan R, Reid K. Essential knowledge for family caregivers: A qualitative study. Int J Palliat Nurs. 2013 Aug 1;19:383–8.
[12] Verbakel E, Metzelthin SF, Kempen GIJM. Caregiving to Older Adults: Determinants of Informal Caregivers' Subjective Well-being and Formal and Informal Support as Alleviating Conditions. J Gerontol B Psychol Sci Soc Sci. 2016 Apr 29;gbw047.
[13] Figma: the collaborative interface design tool. [Internet]. Figma. [cited 2023 Feb 24]. Available from: https://www.figma.com/

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230311

Can Synthetic Images Improve CNN Performance in Wound Image Classification?

Leila MALIHI[a], Ursula HÜBNER[b], Mats L. RICHTER[a], Maurice MOELLEKEN[c], Mareike PRZYSUCHA[b], Dorothee BUSCH[d], Jan HEGGEMANN[e], Guido HAFER[e], Stefan WIEMEYER[e], Gunther HEIDEMANN[a], Joachim DISSEMOND[c], Cornelia ERFURT-BERGE[d], Carlotta BARKHAU[f], Achim HENDRIKS[f] and Jens HÜSERS[b,1]
[a]*Institute of Cognitive Science, Osnabrück University, Germany*
[b]*Health Informatics Research Group, Osnabrück University of AS, Germany*
[c]*Department of Dermatology, Venerology and Allergology, University Hospital of Essen, Germany*
[d]*Department of Dermatology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg, Germany*
[e]*Christian Hospital Melle, Niels Stensen Hospitals, Germany*
[f]*Symbic GmbH, Osnabrück, Germany*
ORCiD ID: Jens Hüsers https://orcid.org/0000-0003-3324-9155

Abstract. For artificial intelligence (AI) based systems to become clinically relevant, they must perform well. Machine Learning (ML) based AI systems require a large amount of labelled training data to achieve this level. In cases of a shortage of such large amounts, Generative Adversarial Networks (GAN) are a standard tool for synthesising artificial training images that can be used to augment the data set. We investigated the quality of synthetic wound images regarding two aspects: (i) improvement of wound-type classification by a Convolutional Neural Network (CNN) and (ii) how realistic such images look to clinical experts (n = 217). Concerning (i), results show a slight classification improvement. However, the connection between classification performance and the size of the artificial data set is still unclear. Regarding (ii), although the GAN could produce highly realistic images, the clinical experts took them for real in only 31% of the cases. It can be concluded that image quality may play a more significant role than data size in improving the CNN-based classification result.

Keywords. wound imaging, data augmentation, convolutional neural network, classification, artificial intelligence, generative adversarial networks, synthetic images

1. Introduction

Artificial intelligence (AI) systems can support health professionals in wound care by automatically recognising wound characteristics such as maceration [1] and infection [2] in wound images, thereby helping to standardise documentation and to curtail record-

[1] Corresponding Author: Jens Hüsers, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: j.huesers@hs-osnabrueck.de.

keeping efforts. However, AI-based systems must perform at a very high level to become clinically relevant. Typically, Machine Learning (ML) based AI systems require a large amount of labelled training data to achieve this level, in particular when the complexity of the domain requires dense coverage. Such amounts of data are sometimes difficult to obtain for secondary use in healthcare, where data access and processing depend on the patient's consent. To use ML for rather sparse data, augmentation techniques can artificially inflate the data basis. Basic augmentation of image data sets can be achieved simply by standard transforms like randomly shifting, rotating, and mirroring the raw image. Beyond these simple methods, ML-based computer vision systems can learn how to generate authentic images of medical entities that do not exist in reality [3]. These systems have seen substantial development in recent years [4], e.g., in a study on ophthalmic images, an AI image generator [5,6] could provide synthetic training images that improved ML-based classification.

We transferred this idea to the domain of chronic wounds, where taking wound images is a standard procedure, but their availability for secondary use can be problematic. To augment training data for a classification task by artificial wound images, generative adversarial networks (GAN) can be employed. However, it is unclear if the desired effect on the classification task materialises and if the artificial images resemble real wound images. We, therefore, investigated the quality of such images regarding two questions: Do they improve the training of a convolutional neural network (CNN)? And second: Do they look realistic to human experts?

2. Methods

Two specialised wound care facilities in Germany, the Christian Hospital Melle and the Department of Dermatology, Venerology and Allergology of the University Hospital Essen, provided wound images taken in routine wound care showing two distinct wound types: diabetic foot ulcers and venous leg ulcers. The information on the wound type and the images were retrieved from the patient records. In total, 987 images were curated to build the dataset - 480 images of diabetic foot ulcers and 507 of venous leg ulcers. The average raw image resolution is 2705 x 3374 pixels. A clinician located the wound in the image with a bounding box used to crop the wound with an additional margin of 75 pixels. These cropped images were scaled to 256 by 256 pixels and finally checked for any errors by a second clinician. Next, this dataset was randomly split aiming for a ratio of 9:1 into a training set containing 864 images (88 %) and a hold-out test set containing 123 images (12 %).

A GAN of the StyleGAN3 architecture was trained to produce colorized synthetic but realistic looking wound images using the curated dataset described above. As with any GAN, StyleGAN3 comprises two neural networks. A first generator network produces natural-looking images, hereby trying to fool a second network, the discriminator, whose task is to distinguish the artificial from the real-world images. The discriminator's decision amplifies the generator's learning process and helps to improve the quality of the generated images. Simultaneously, the discriminator improves its performance on the increasing difficulty of the task as artificial images become ever more realistic [3]. We used a RTX8000 GPU for the training process.

In the next step, we trained deep CNN based on the Xception architecture [7]. As a basic data augmentation method, we randomly manipulated the training images' brightness and shear to account for the unstandardised lighting conditions and viewing

angles in the original wound images. The models were trained on a maximum of 100 epochs with an early stopping callback of 20 epochs of non-improvement. The ones with the lowest validation loss were evaluated on the test set using accuracy, recall, precision, and F1-score as performance indicators. The model training was done in Python using Tensorflow 2.9 with an NVIDIA Tesla T4 GPU. In the following experiment, we used this setup to evaluate the effect of augmentation by GAN-generated images on model performance (first research question): First, we trained a neural network using only the real images (plus the basic augmentations). Then, we doubled and quadrupled the dataset size by including the synthetic images.

To assess the second research question, we recruited 217 clinicians with self-reported wound expertise of 5 (median) on a scale from 1 (no experience) to 7 (maximum experience). They were asked to identify the synthetic images among a random subset of 60 real-world images from the test set and a random subset of 64 generated images in an online survey. The survey data was analysed using contingency tables comparing the ground truth (real vs synthetic) and the expert's predictions.

3. Results

We trained a GAN that produced synthetic wound images for diabetic foot ulcers and venous leg ulcers and performed the experiment that resulted in three classification models (Table 1). All models showed convergence and early stopping triggered before the maximum of 100 epochs in all training runs. All models yielded acceptable performance metrics. With the growing dataset size (that we obtained by augmenting the original training dataset with synthetic images), the accuracy and the F1-score metrics improved steadily with the dataset size (Table 1). However, precision and recall did not follow this trend. The maximum precision value was achieved with the largest dataset; however, the second largest dataset produced a precision value smaller than that of the original dataset. In contrast, the maximum recall value was achieved with the second largest dataset. However, the F1-score increased. Generally, the effect of the dataset's size on performance was inconsistent, depending on the performance indicator.

Table 1. Performance of the deep neural network classifier trained with three different training sets (raw, doubled and quadrupled dataset size, see left column). All three models with different training sets were evaluated on the same hold-out set.

Training Data	Size	Accuracy	Precision	Recall	F1
Original raw images	864	0.851	0.872	0.850	0.844
Double dataset size (50% real, 50% synthetic)	1,728	0.870	0.852	0.881	0.867
Quadrupled dataset size (25% real, 75% synthetic)	3,456	0.878	0.885	0.871	0.878

The GAN's performance on the quality of the images (research question ii) was investigated based on 26,908 decisions made by the 217 clinicians, each evaluating 124 (60 real and 64 synthetic) images. Table 2 shows the 26,908 decisions tabulated against the ground truth. The upper contingency table (table 2) reveals that out of the 13,020 decisions on the synthetic images, the clinicians regarded the images in 4,061 cases as real ones (31%).

Similarly, among the 12,321 decisions of voting for "real", there were 4,061 decisions that were based on synthetic images (33%). These conditional frequencies showed that clinicians regarded synthetic images in one-third or less of their decisions as real ones. Overall, 64% of the decisions (17,219) regarding synthetic and real images

were correct. Conversely, there were also real images regarded as synthetic by the clinicians.

Figure 1 provides epitomic images of synthetic and real images and the majority vote of the clinicians. As figure 1 demonstrates, "bad" synthetic images were identified as such (upper left) by many clinicians. Still, there were also "good" synthetic images, as the left lower corner example reveals. Likewise, true real images, such as the slightly blurred one in the upper right corner of Figure 1, were rated as synthetic.

Table 2 Contingency tables with conditional frequencies of ground truth contrasted by clinical decision

Aggregation		**Ground Truth**		
Per columns		Synthetic	Real	Total
Decision	Synthetic	69% (8,959)	41% (5,628)	54% (14,587)
	Real	**31% (4,061)**	59% (8,260)	46% (12,321)
	Total	100% (13,020)	100% (13,888)	100% (26,908)
Per rows		Synthetic	Real	Total
Decision	Synthetic	61% (8,959)	39% (5,628)	100% (14,587)
	Real	**33% (4,061)**	67% (8,260)	100% (12,321)
	Total	48% (13,020)	52% (13,888)	100% (26,908)

4. Discussion

In this study, a GAN was trained to produce synthetic wound images to augment a dataset for CNN training. The resulting CNN classification models showed good performance (F1-score ≥ 0.844) and tended to improve slightly with growing dataset size concerning F1-score and accuracy, albeit not for recall and precision. This small and inconsistent effect of data size does not follow previous studies, which showed that CNNs generally improve with more data available [8]. A possible reason for this finding could be the quality of the synthetic images. The clinicians' decisions provided valuable information on how well the GAN-generated images match real images. When presented with a synthetic image, the clinicians took them for real only in 31% of the cases. However, the GAN was capable of producing high-quality synthetic wound images like the one in the lower left panel of Figure 1.

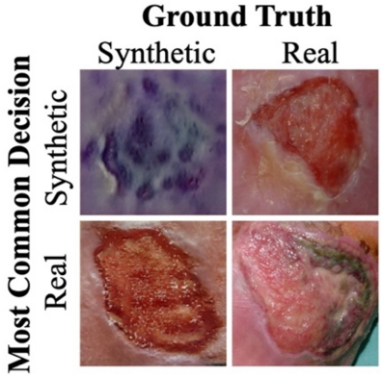


Figure 1. Image examples illustrating the contingency tables (Tab.2)

While it seems possible to improve the GAN to create better synthetic images, we assume that another direction of future research is more promising: If not the complete

set of synthetic images was used for training but only the ones that appeared realistic to clinicians, performance of the CNN is likely to rise with data size. Though this procedure requires manual classification into "good" and "bad" synthetic images, the effort is still lower compared to the acquisition of a larger database of real images.

Additional information obtained from clinicians' comments on the synthetic images might lead to further improvements: They frequently commented on unfavourable lighting conditions and image resolution that impeded identification. These comments were probably provoked by the unstandardised wound images we used in our curated dataset. We thus anticipate that wound images standardised with respect to angle, distance, lighting conditions, and proportion of wound area in the image would improve the image generation by GANs. Furthermore, the health professionals commented, that the wound images lacked peri-wound context information (due to a-priori cropping). We assume that this missing context may have made the task difficult for the clinicians as it differed too much from a real-world wound imaging setting. To close the gap to real world-wound imaging, GANs should aim additionally to synthesise peri-wound characteristics. It remains to be tested if the peri-wound context would provide useful additional features for CNN classification. Considering these options, we are optimistic about bringing GAN-generated images to a level of quality that will significantly improve automated classification from sparse training sets.

Acknowledgement

This study is part of a project funded by the German Ministry of Education and Research (BMBF Grant No. 16SV8616). We thank our project partners of the University Osnabrück New Public Health and apenio GmbH for their support and cooperation. This work was supported by a fellowship within the IFI program of the German Academic Exchange Service (DAAD).

References

[1] Hüsers J, Hafer G, et al. Automatic Classification of Diabetic Foot Ulcer Images: A Transfer-Learning Approach to Detect Wound Maceration. IOS Press; 2022. doi:10.3233/SHTI210919.

[2] Yogapriya J, Chandran V, Sumithra MG, Elakkiya B, Shamila Ebenezer A, Suresh Gnana Dhas C. Automated Detection of Infection in Diabetic Foot Ulcer Images Using Convolutional Neural Network. Journal of Healthcare Engineering. 2022 Apr 6;2022. doi:10.1155/2022/2349849.

[3] Yi X, Walia E, Babyn P. Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by wasserstein distance for dermoscopy image classification. arXiv preprint arXiv:1804.03700. 2018 Apr 10.. doi:10.48550/ARXIV.1804.03700.

[4] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 (pp. 10684-10695). doi:10.48550/ARXIV.2112.10752.

[5] Kim M, Kim YN, Jang M, Hwang J, Kim HK, Yoon SC, Kim YJ, Kim N. Synthesizing realistic high-resolution retina image by style-based generative adversarial network and its utilization. Scientific Reports. 2022 Oct 15;12(1):17307. doi:10.1038/s41598-022-20698-3.

[6] Chen JS, Coyner AS, Chan RP, Hartnett ME, Moshfeghi DM, Owen LA, Kalpathy-Cramer J, Chiang MF, Campbell JP. Deepfakes in ophthalmology: applications and realism of synthetic retinal images from generative adversarial networks. Ophthalmology Science. 2021 Dec 1;1(4):100079. doi:10.1016/j.xops.2021.100079.

[7] Chollet F. Xception: Deep learning with depthwise separable convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1251-1258). doi:10.1109/CVPR.2017.195.

[8] Luo C, Li X, Wang L, He J, Li D, Zhou J. How does the data set affect cnn-based image classification performance?. In2018 5th international conference on systems and informatics (ICSAI) 2018 Nov 10 (pp. 361-366). IEEE. doi:10.1109/ICSAI.2018.8599448.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230312

Few-Shot Meta-Learning for Recognizing Facial Phenotypes of Genetic Disorders

Ömer SÜMER[a,1], Fabio HELLMANN[a], Alexander HUSTINX[b], Tzung-Chien HSIEH[b], Elisabeth ANDRÉ[a] and Peter KRAWITZ[b]
[a] *Chair for Human-Centered Artificial Intelligence, University of Augsburg, Germany*
[b] *Institute for Genomic Statistics and Bioinformatics, University of Bonn, Germany*

Abstract. Computer vision has useful applications in precision medicine and recognizing facial phenotypes of genetic disorders is one of them. Many genetic disorders are known to affect faces' visual appearance and geometry. Automated classification and similarity retrieval aid physicians in decision-making to diagnose possible genetic conditions as early as possible. Previous work has addressed the problem as a classification problem; however, the sparse label distribution, having few labeled samples, and huge class imbalances across categories make representation learning and generalization harder. In this study, we used a facial recognition model trained on a large corpus of healthy individuals as a pre-task and transferred it to facial phenotype recognition. Furthermore, we created simple baselines of few-shot meta-learning methods to improve our base feature descriptor. Our quantitative results on GestaltMatcher Database (GMDB) show that our CNN baseline surpasses previous works, including GestaltMatcher, and few-shot meta-learning strategies improve retrieval performance in frequent and rare classes.

Keywords. Facial genetics, rare genetic disorders, image analysis, few-shot learning, meta-learning, imbalanced data, deep learning.

1. Introduction

Genetic disorders affect more than 5% of the population [1]; however, physicians might fail to spot and clinically diagnose most of them. There is a set of genetic conditions, and 30-40% of them are known to affect craniofacial development and facial morphology [2], and computer vision can help recognize skull alterations from facial images [3]. The output of such a system can support physicians in diagnosing rare syndromes and eventually lead to therapeutic interventions.

Previous literature uses geometric information, facial landmarks, and handcrafted features around face regions [4], however, a small number of subjects and syndromes limit their use in clinical settings. Shukla et al. [5] combined convolutional neural network features in face regions and used SVM classifiers. Recent studies [6,7] showed that end-to-end deep learning-based methods could substantially improve facial phenotyping.

Still, the number of samples in real-life situations and databases shows considerable variation across disorders. This makes training deep convolutional networks not feasible, as in any object classification task. The nature of the problem necessitates addressing

[1] Corresponding Author: Ömer Sümer, E-mail: oemer.suemer@informatik.uni-augsburg.de.

data imbalance and few-shot classification in facial phenotype analysis. Collecting facial images of rare facial genetic disorders requires lots of effort. Most of the previous works do not have publicly available databases to benchmark computer vision methodologies. GestaltMatcher Database[2] [7] is a recent effort to carry automated facial phenotyping forward. This paper presents a deep learning baseline that depends on a better facial recognition model and a few-shot meta-learning approach for unseen facial genetic disorders based on a highly imbalanced distribution of disorders.

2. Method

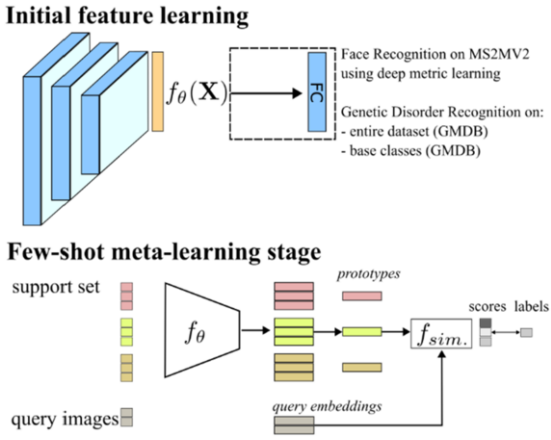


Figure 1. Workflow of initial feature learning and few-shot meta-learning: top) the initial feature learning is done either on face recognition task or genetic disorder classification; bottom) the learned representation is used in the few-shot meta-learning stage.

Figure 1 depicts the workflow of our proposed approach for facial phenotype recognition. The initial step in facial phenotype learning is to learn a solid initial representation. We trained a convolutional neural network backbone for this task by adopting the metric learning-based Arcface loss [8] in face recognition. Subsequently, the few-shot meta-learning stage aims to learn facial phenotypes from highly imbalanced data where most categories have limited samples.

There are separate support and query sets to learn to compare in the training and testing phases. These sets are created in an episodic manner. The bottom part of Figure 1 takes sampled episodes of support and query images, and first extracts features using the backbone encoder by initializing from face recognition pre-trained weights. The few-shot learning is defined according to the number of categories (N) and samples (K) in each support group in the support set. The task is described as K-way N-shot in previous literature [9]. During the training, the centroid of each embedding vector per class c is calculated: $c_k = \frac{1}{S_k}\sum_{x_i,y_i \in S_k} f_\Theta(x_i)$ where x_i and y_i are images and corresponding labels in each group of support set, S_k.

Furthermore, differing from [9], each episode has several K-way N-shot tasks. It refers to predicting the category of a query sample from K classes or N examples per

[2] It is accessible for clinicians and computer scientists under the following link: https://db.gestaltmatcher.org/

class in the support set. This setting learns a feature embedding that can retrieve samples belonging to the same category using a similarity metric. The main difference here is that meta-learning is independent of the tasks and can better generalize on unseen classes. In Prototypical Networks [9], the distance (or similarity) function is Euclidean distance. However, previous literature in facial phenotype recognition [7] used cosine similarity for the retrieval task. To make our meta-training as compatible as possible with our end task, we used cosine similarity between query embeddings, $f(x_i)$ and class centroids, c_k, and calculated logits as follows:

$$p(y = k \mid i) = \frac{e^{\tau \cdot \langle f_\Theta(x_i), c_k \rangle}}{\sum_k' e^{\tau \cdot \langle f_\Theta(x_i), c_k \rangle}} \quad (1)$$

where τ is a learnable scalar that we applied to scale the values before applying the Softmax function following the related literature [10,11].

We conducted our experiments on version 1.0.3 of the GMDB [7]. In version 1.0.3, the database contains 7.459 images of 449 syndromes. In addition to training and validation sets, there are two separate galleries and test sets for frequent and rare disorders. In both, faces are detected and aligned by RetinaFace [8]. Using five facial key points, we performed 5-point similarity alignment and normalized faces to the size of 112x112 pixels. During the training of baseline classification and few-shot meta-learning models, we only applied channel mean and standard deviation normalization according to the train set statistics and random horizontal flipping. When training both whole-set classification and few-shot meta-training, we used an SGD solver with a constant learning rate of 0.001 and weight decay of 0.0005 for 25 epochs. We used validation retrieval performance, specifically, the nearest neighbor retrieval of validation samples' feature embeddings to all training sets for model selection. During the few-shot meta-training, we sampled each episode containing four tasks, and the total number of episodes was kept at 100 and trained for 25 epochs.

Using the 512-dimensional embedding vector as feature representation, we evaluated the performance of our classification and nearest-neighbor approach in terms of top-k accuracies in the frequent and rare test sets in the GMDB. Following Hsieh et al. [7], learned facial embeddings were evaluated using two settings as follows:

1. The retrieval task reports top-k accuracies using k-nearest neighbors based on feature embeddings and cosine distances from the frequent gallery and frequent test sets.
2. The retrieval task reports top-k accuracies using k-nearest neighbors based on feature embeddings and cosine distances from the 10-Fold Cross-Validation rare gallery and rare test sets.

In all experiments, we calculate accuracies for Top-{1, 5, 10, 30} retrieval. We also reported the classification performance of [7] that reports only top-k accuracies using softmax outputs based on the frequent test set.

3. Results

Table 1 depicts the results of our ablation study. As we aim to improve the retrieval performance on both tasks, we only evaluated GestaltMatcher DCNN using predictions trained with cross-entropy loss. The performance of GestaltMatcher DCNN trained on v1.0.3 of the database is aligned with the published results in [7]. Top-1 accuracy varies in the ranges of 15% to 21% in frequent and rare sets where the total number of classes

is 204 and 245, respectively. Our stronger baseline, a ResNet-50 trained on MS1MV2 using ArcFace loss (Enc-healthy), performed 34.06% top-1 accuracy in the frequent set, whereas GestalthMatcher DCNN's retrieval performance remains at 15.96%.

Few-shot meta baseline that we adopted in our experiments is a 10-way 3-shot task with 2 query samples in each task. Following [11], we sampled multiple tasks in each episode. The reported experiments are done using four tasks per episode. Table 2 shows the retrieval performance of few-shot meta-learning models on both frequent and rare test sets.

Table 1. Performance comparison of GestaltMatcher DCNN and our baseline models on GMDB (v1.0.3).

Method	Top-1	Top-5	Top-10	Top-30
Frequent Set				
GestaltMatcher DCNN (7)				
Classification	21.21	42.08	54.60	73.92
Retrieval	15.96	33.83	45.46	69.64
Enc-healthy	34.06	53.96	64.42	**81.28**
Enc-all (GMDB)	**42.50**	58.18	65.26	78.08
Enc-base (GMDB)	40.47	**60.71**	**67.29**	79.09
Rare Set				
GestaltMatcher DCNN (7)				
Retrieval	19.26	36.28	44.07	60.73
Enc-healthy	26.31	42.62	46.98	62.92
Enc-all (GMDB)	26.40	42.36	50.42	65.76
Enc-base (GMDB)	**28.25**	**44.88**	**52.00**	**66.18**

Table 2. Few-shot meta baseline and feature-level fusion on GMDB (v1.0.3) retrieval task.

Method	Top-1	Top-5	Top-10	Top-30
Frequent Set				
GMDB-fs	48.06	68.13	75.89	85.67
(feature-level fusion)				
+Enc-healthy	47.55	**68.47**	**77.23**	**88.69**
+Enc-all (GMDB)	**47.55**	67.62	74.20	84.65
+Enc-base (GMDB)	47.22	67.96	74.71	84.82
Rare Set				
GMDB-fs	30.21	48.19	56.39	71.07
(feature-level fusion)				
+Enc-healthy	32.89	**50.65**	**57.89**	**71.39**
+Enc-all (GMDB)	30.88	48.29	56.57	70.54
+Enc-base (GMDB)	**33.08**	48.37	56.65	70.72

Few-shot meta-training (GMDB-fs) improves the top-1 frequent test accuracy of the best GMDB-trained baseline models, Enc-all and Enc-base by 7.59%, and 5.56%, respectively. This improvement is not limited to top-1 retrieval, it is also retained in different neighbor retrieval. We initialized GMDB-fs models using healthy encoding.

Table 3. Comparison of n categories with 4 tasks per episode and 10 categories with n-shot and n-query.

	Frequent				Rare			
	Top-1	Top-5	Top-10	Top-30	Top-1	Top-5	Top-10	Top-30
n-categories 3-shot / 2-query								
5	**49.24**	66.44	75.04	84.49	27.70	54.41	54.44	69.37
10	48.06	**68.13**	**75.89**	85.67	30.21	48.19	56.39	71.07
15	47.89	67.96	75.72	**86.51**	**31.63**	**49.35**	**58.17**	**72.95**
20	48.06	67.62	74.37	84.65	27.76	47.04	55.29	69.33
n-shot/n-query, 10-categories								
1/4	44.35	65.94	73.19	84.65	29.37	46.99	56.26	69.47
2/3	44.35	67.12	74.87	**86.34**	**31.77**	**49.26**	**57.66**	71.04
3/2	47.05	**69.14**	**76.05**	**86.34**	30.15	48.38	57.17	**71.34**
4/1	**48.23**	68.13	75.21	84.65	27.76	45.79	55.58	68.72

In both frequent and rare sets, feature-level fusion with the healthy encoder performed the best in nearly all retrieval tasks. In top-1 rare retrieval, fusion with Enc-base gives the best accuracy, 33.08%. Even though Enc-all and Enc-base perform better than Enc-healthy, their performance on feature fusion is limited.

4. Discussion

We observed differences in the model's behavior when evaluating the few-shot meta-based training with different sets of configurations (Table 3). These variables affect the difficulty of few-shot tasks and must be examined in depth. One is the number of ways

(categories) to define possible classes in a support set. The best retrieval performance is in the 10 and 15 categories. We consider this behavior related to the complexity of classification tasks in each episode. We picked the 10-way to evaluate other parameters that affect the performance of episodic training. These are the number of images in each class in the support set (k-shot) and the number of query images. Table 3 (bottom) presents evaluation performance using the different number of shots and queries. A higher number of shots improves frequent set retrieval performance; however, the 4-shot setting performs worse in rare set retrieval. The minimum number of shots seems descriptive enough according to the n-way task learned. We could not increase n-shot and queries as the minimum number of samples per class in the training set was 5. The best overall performance is in 2-shot or 3-shot settings.

5. Conclusion

In this study, we trained a state-of-the-art face recognition model on standard face recognition databases. We transferred learned representations to our low-resource target data domain for facial phenotype recognition for genetic disorders. We addressed the issue of data scarcity and imbalanced data using a few-shot meta-based learning approach. This improved genetic disorder recognition of unseen genetic conditions compared to the recently published GestaltMatcher DCNN; however, our study has certain limitations. We need more samples of rare diseases to ensure a fine-grained analysis and a user study of how AI models and clinicians' decisions deviate. In future work, using generative models on either image or feature level, synthesized samples can also be added to few-shot training and reduce the effect of uneven class distribution.

References

[1] Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. American journal of human genetics. 1988; 42(5): 677-693.
[2] Hart TC, Hart PS. Genetic studies of craniofacial anomalies: clinical implications and applications. Orthodontics & craniofacial research. 2009 February; 12(3): 212-220.
[3] Ferry , Steinberg J, Webber C, FitzPatrick DR, Ponting CP, Zisserman A, et al. Diagnostically relevant facial gestalt information from ordinary photos. eLife. 2014 June.
[4] Loos HS, Wieczorek D, Würtz RP, Malsburg Cvd, Horsthemke B. Computer-based recognition of dysmorphic faces. European Journal of Human Genetics. 2003; 11(8): 555-560.
[5] Shukla P, et al. A deep learning framework for recognizing developmental disorders. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV); 2017: IEEE. p. 705-714.
[6] Gurovich Y, Hanani Y, Bar O, Guy N, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nature medicine. 2019; 25(1): 60-64.
[7] Hsieh TC, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nature genetics. 2022 February; 54(3): 349-357.
[8] Deng J, Guo J, Xue N, Zafeiriou S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. p. 4690-4699.
[9] Snell J, Swersky K, Richard Z. Prototypical networks for few-shot learning. In Advances in neural information processing systems; 2017: Curran Associates, Inc.
[10] Gidaris S, Komodakis N. Dynamic few-shot visual learning without forgetting. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 4367-4375.
[11] Yinbo C, Liu Z, Xu H, Darrell T, Wang X. Meta-baseline: Exploring simple meta-learning for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 9062.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230313

Model-Driven Dementia Prevention and Intervention Platform

Dr. Markus SCHINLE[a,1], Michael DIETRICH[a], Dr. Simon STOCK[b], Marius GERDES[b] and Prof. Wilhelm STORK[b]
[a] *FZI Research Center for Information Technologies, Germany*
[b] *KIT Karlsruhe Institute of Technology, Germany*

Abstract. Most types of dementia, including Alzheimer's disease, are not curable. However, there are risk factors, such as obesity or hypertension, that can promote the development of dementia. Holistic treatment of these risk factors can prevent the onset of dementia or delay it in its early stages. To support individualized treatment of risk factors in dementia, this paper presents a model-driven digital platform. It enables monitoring of biomarkers using smart devices from the internet of medical things (IoMT) for the target group. The collected data from such devices can be used to optimize and adjust treatment in a patient in the loop manner. To this end, providers such as Google Fit and Withings have been connected to the platform as example data sources. To achieve treatment and monitoring data interoperability with existing medical systems, internationally accepted standards such as FHIR are used. The configuration and control of the personalized treatment processes are achieved using a self-developed domain-specific language. For this language, an associated diagram editor was implemented, which allows the management of the treatment processes through graphical models. This graphical representation should help treatment providers to understand and manage these processes more easily. To investigate this hypothesis, a usability study was conducted with twelve participants. We were able to show that such graphical representations provide advantages in clarity in reviewing the system, but lack in easy set-up (compared to wizard-style systems).

Keywords. model-driven intervention planning, internet of medical things, digital prevention, treatment platform, precision care, Alzheimer's disease, IoMT, FHIR

1. Introduction and Motivation

Currently, dementia is not curable, and treatments are aimed at improving the quality of life of those affected, for example, by treating certain symptoms with medication [1]. Therefore, the prevention of dementia is of particular interest. The most common form of dementia, Alzheimer's disease, is favored by modifiable risk factors. These include treatable diseases such as diabetes and lifestyle-related factors such as an unhealthy diet, lack of physical activity, or sleep disturbances [2]. Personalized interventions can help to preserve the mental abilities of people at-risk or, in the early stages of dementia, delay the onset and progression of the disease. This involves combining different treatment approaches based on individual risk factors, such as nutritional counseling, exercise therapy, or drug treatment [3]. Initial success in decreasing the risk for the development

[1] Corresponding Author: Dr. Markus Schinle, E-mail: schinle@fzi.de.

of dementia was achieved in the FINGER study [4]. In this study, people at increased risk of dementia were treated individually over two years with nutrition therapy, exercise therapy, and memory training [5]. This treatment showed promise to maintain or improve cognitive function. One challenge here is that the treatments for prevention or intervention in dementia take place over several years and require continuous adaptation to, and monitoring of, the beneficiary individual's state of health and needs. Our previous studies indicated that medical professionals are willing to use digital tools to support this treatment process [6]. Nevertheless, from an ethical point of view, the potential for surveillance and discrimination due to the misuse of such a platform is seen as a threat. On the other hand, the increasing spread of technologies, such as smart fitness wristbands, indicates a greater willingness among the population to use health-related data for themselves.

2. Methods and Definition of Research Questions

Our work aims to develop a model-driven Internet of Medical Things (IoMT) platform to support personalized mHealth applications for dementia prevention and intervention. Besides a mobile application, IoMT devices, such as wearables, will be integrated into the treatment process to monitor individuals' progress, digital biomarkers, and risk factors. IoMT-based personalized treatment may involve the combination of different treatment approaches as well as objectified feedback using multiple heterogeneous devices with different characteristics and sensors. Accordingly, the treatment process is multi-layered and variable. Therefore, the medical professional shall use models to manage and configure the treatment processes. These models should graphically represent and thus facilitate the understanding of the processes as well as the handling of the system for the medical professional. Graphical models are already being used on the Internet of Things (IoT) field to allow non-technical users, e.g. to configure data pipelines to process IoT data streams. In the software development domain, models are used to document and improve understanding of software systems. Model-driven software development uses models of systems for the automatic generation of program code. Here Domain Specific Languages (DSL) are used to express the elements and problems of a domain, thus, a limited knowledge area [7]. To facilitate the use of a platform for medical professionals, a DSL for modeling IoMT-based personalized treatment processes with a focus on holistic dementia prevention is presented in this work. The DSL is useable to specify, manage, and configure the IoMT-supported treatment processes. As a procedure for developing these artifacts, we chose the Design Science Research Method (DSRM) by Peffers et al. [8]. This is broken down within a process into the six steps of problem identification and motivation, definition of goals for the solution, design and development of the solution artifact, demonstration of the solution artifact, evaluation of effectiveness and efficiency, and communication. This enables a systematic development of artifacts for an identified problem, as well as their subsequent evaluation based on the findings of literature research and the deduction of requirements of such a platform. Therein, the following questions are addressed:

1. How must a DSL be designed to specify an IoMT-supported treatment process?
2. How can models be created using the DSL in a model-driven software platform?

3. Results

Addressing risk factors has been shown as an iterative process. Initially, a diagnosis is made by a medical professional. During this process, an individual is examined for risk factors such as lack of exercise, obesity, or chronic diseases like high blood pressure. Based on these risk factors, the medical professional then decides on appropriate interventions to reduce these risk factors. For example, increasing physical activity through sports therapy can manage a lack of exercise. To objectively record the treatment progress, the individual's treatment or risk factors are continuously monitored. The medical professional selects suitable IoMT devices and questionnaires and issues them to the individual. These artifacts then collect the required health data. Based on this data, the medical professional can continuously adjust and improve treatment or identify new risk factors. For this purpose, our DSL requires at least one concrete syntax to represent the elements of this abstract syntax. In our platform, it is implemented in the form of textual forms and graphical diagrams. Therefore, a treatment plan's elements are e.g., represented in a diagram. The structure of a treatment plan diagram is shown in Fig. 1. Treated risk factors, interventions, and challenges used in the process, as well as devices, questionnaires, and notifications to remind the individual, should be represented here as nodes. To build the relationship between the nodes and to compose the treatment plan, these nodes shall be connected by edges (arrows). A connection from a risk factor to an intervention means that the connected intervention treats the risk factor. Interventions are connected to so-called Challenges, each of which is a component of the intervention. The outgoing edges of a challenge describe which Devices or Questionnaires are used for monitoring as part of the challenge and which Reminder for that challenge have been created for the individual.

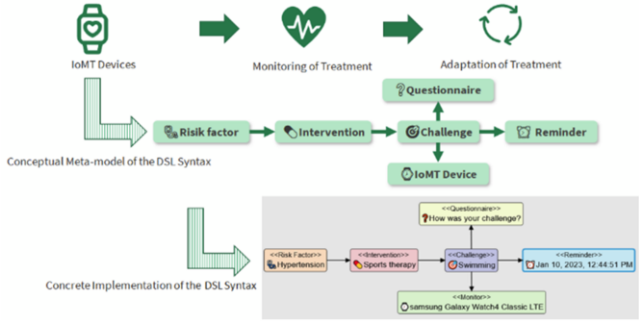


Figure 1. The design steps of the DSL are shown in three abstraction levels. On the top, the overall treatment process is displayed. In the middle, a conceptual representation of the DSL is presented. At the bottom, our system implementation is shown.

A system architecture shown in Fig. 2 was designed to use this DSL within the context of a digital dementia prevention and intervention platform. An Angular-based web application for medical professionals is implemented to interact with the platform [9]. This allows them to view and manage the treated individuals. The JSONForms framework displays and manages elements such as individuals, treatment plans, devices, and the elements in the settings tab [10]. The *Treatment Service* handles the management of individuals, their diagnosis, and treatment. This service was implemented based on technologies of the Eclipse Modelling Framework (EMF) [11]. The Eclipse Graphical Server Language Platform (GLSP) was used to implement the diagram editor for treatment plans [12]. This enables the development of web-based diagram editors consisting of a client and an associated server. The GLSP client was integrated into the

Angular web application and the GLSP server into the Treatment Service. The *Monitoring Service* enables retrieval and collection of monitoring data from individuals' IoMT devices A Nest server was used to implement this service. For data storage, the microservice also uses MongoDB database. To fulfill the requirement of interoperability of the platform with existing medical infrastructure, the Fast Healthcare Interoperability Resources (FHIR) standard was used to persist treatment data [13]. This defines a uniform data format and interfaces for exchanging health data between software systems. For some elements, mapping to FHIR resources is not straightforward. This includes, e.g., additional information on risk factors. There is no suitable resource for these. Therefore, the condition information of a risk factor is bundled in the Basic resource. The Basic resource is used for concepts that are not yet included in the standard. To store information about risk factors, the Basic resource is complemented by its own profile. A separate profile was also defined for other resources of the FHIR standard to adapt them to represent the data of the platform. These profiles define constraints on existing fields of the resource, as well as extensions to the resource. It can be used to persist the data within the FHIR Service. Individuals can eventually interact with medical professionals using a cross-platform mobile application.

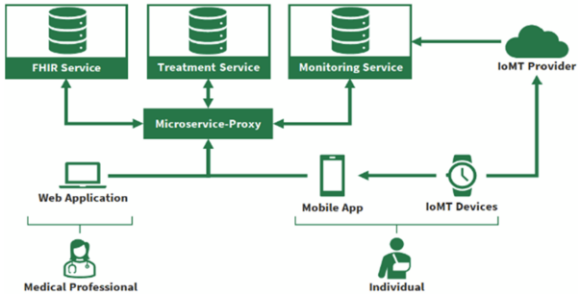


Figure 2. Architecture of the risk factor intervention platform. A microservice proxy and IoMT service providers enable communication between the patient/individual side and the medical professional.

4. Evaluation and Discussion

For the evaluation of the proposed DSL, the platform was checked for meeting all defined requirements. In addition, a usability study was conducted with a 20-minute questionnaire-based survey. In this, a one-sample treatment scenario was to be modeled by the subjects using the developed DSL as well as another variant of the platform that allows treatment to be administered via forms only (wizard-style). A total of 12 subjects participated in the study, ranging from 10 to 25 years of computer experience. For the quantitative assessment of usability, the usage questionnaire by Prümper was used, which is based on ISO 9241-110 standard [14]. Overall, administration via forms performed better than administration via the developed DSL. The results of the user questionnaire showed the advantages of forms over models, especially in terms of ease of learning and conformity to expectations. The model-driven platform, however, performed better in fault tolerance and self-descriptiveness. Ten of the twelve participants also indicated a preference for form-based over model-based management. The primary reason for this is that the latter enables easier use without prior knowledge due to the predefined structure. The advantages of model-based management were located by the participants in the clarity and reviewability of the system. The study could

not confirm the hypothesis of improved usability using models, at least for the selected example scenario. The selected scenario was, however, limited by the functional scope of the form-based variant of the platform and the time frame of the evaluation. For the triage and understanding of treatment plans with more elements, the graphical models could nevertheless be advantageous.

5. Conclusion and Future Work

Based on the findings of the literature research and the defined requirements of the platform, a DSL was designed, which supports the management of treatment plans and the configuration of the monitoring. For this purpose, the syntax of the DSL was designed to create models for the planning and monitoring of treatments. Furthermore, the treatment plan can also be used to configure the monitoring of the treatment. The individual's equipment and questionnaires can be selected and added to the treatment plan. For devices, it is also possible to select which parameters, such as activity or sleep, are to be recorded. A transformation rule was developed that maps these models to corresponding FHIR resources. In the future, these FHIR resources will be retrieved and displayed by the individual's mobile application. Therein, the individual should be able to display an overview of the active treatment plans and their challenges, i.e., goals and tasks to be completed. We want to deliver reminders to the individual via the app in the form of notifications so that they are reminded of their challenges and interventions. This work presented a DSL and data representation, the cornerstone for an overall system capable of dementia risk management and treatment. We are confident in delivering and evaluating user-centered mobile applications soon.

References

[1] Livingston et al., “Dementia prevention, intervention, and care”, Lancet, 390(10113):2673-4, 2017.
[2] George A. Edwards Iii et al., “Modifiable Risk Factors for Alzheimer’s Disease“, doi:10.3389/fnagi.2019.00146, 2019.
[3] Richard S. Isaacson u. a. “The clinical practice of risk reduction for Alzheimer’s disease: A precision medicine approach“, doi: 10.1016/j.jalz.2018.08.004, 2018.
[4] Miia Kivipelto et al., “The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER): study design and progress“, doi: 10.1016/j.jalz.2012.09.012. 2013.
[5] Ngandu et al., “A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial“, doi: 10.1016/S0140-6736(15)60461-5, 2015.
[6] Schinle et al., “Digital Health Apps in the Context of Dementia: Questionnaire Study to Assess the Likelihood of Use Among Physicians”, doi: 10.2196/35961, 2022.
[7] Thomas Stahl, “Modellgetriebene Softwareentwicklung: Techniken, Engineering, Management”, isbn: 9783898648813, url: http://swb.eblib.com/patron/FullRecord.aspx?p=1049742, 2012.
[8] Ken Peffers et al., “A Design Science Research Methodology for Information Systems Research”, doi: 10.2753/MIS0742-1222240302, 2007.
[9] Angular, https://angular.io/, 10.01.2023.
[10] JSONForms, https://jsonforms.io/, 10.01.2023.
[11] Eclipse Modelling Framework Cloud, https://www.eclipse.org/emfcloud/, 10.01.2023.
[12] Graphical Language Server Platform, https://www.eclipse.org/glsp/, 10.01.2023.
[13] Fast Healthcare Interoperability Resources, HL7FHIR, http://hl7.org/fhir, 10.01.2023.
[14] Jochen Prümper, “Fragebogen ISONORM 9241/110-S: Beurteilung von Software auf Grundlage der Internationalen Ergonomie-Norm DIN EN ISO 9241-110”, url: http://www.uselab.tu-berlin.de/wiki/images/6/62/ISONorm_Kurzversion.pdf, 10.01.2023.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230314

The Future of Online Video Consultations in Primary Care: A Qualitative Study

Maria HÄGGLUND[a, b,1], Nadia DAVOODY[b], Anna KRISTENSSON EKWALL[c] and Nasim FARROKHNIA[d]
[a] *Department of Women's and Children's Health, Uppsala University, Sweden*
[b] *Health Informatics Center, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Sweden*
[c] *Department of Health Sciences, Lund University, Sweden*
[d] *Department of Clinical Science and Education, Karolinska Institutet, Sweden*
ORCiD ID: Maria Hägglund https://orcid.org/0000-0002-6839-3651
ORCiD ID: Nadia Davoody https://orcid.org/0000-0003-1813-8529
ORCiD ID: Anna Kristensson Ekwall https://orcid.org/0000-0002-8768-2467
ORCiD ID: Nasim Farrokhnia https://orcid.org/0000-0002-2064-137X

Abstract. The COVID-19 pandemic has significantly increased the use of remote services such as video consultations (VCs). In Sweden, private healthcare providers offering VCs have grown substantially since 2016 and have been controversial. Few studies have focused on physicians' experiences of providing care in this context. Our overall aim was to study physicians' experiences of VCs, here focusing on their suggestions for future improvements. Twenty-two semi-structured interviews were performed with physicians working for an online healthcare provider in Sweden, and analyzed through inductive content analysis. Two themes emerged related to desired future improvements of VCs; blended care and technical innovation.

Keywords. Video Consultations, Telemedicine, Primary Care

1. Introduction

In the wake of the COVID-19 pandemic, telemedicine has become a necessity in healthcare [1,2]. OECD defines telemedicine as "the use of ICT to deliver health care (clinical services only) at a distance" [3], with three subgroups;

(1) telemonitoring, or remote patient monitoring, tele homecare, i.e. use of information and communication technologies (ICTs) to monitor health status at a distance,
(2) store and forward, i.e. an encounter or consult aided by asynchronous transmission of clinical data, and
(3) interactive telemedicine, or video consultations (VCs), real-time teleconsultations, virtual visits, i.e. synchronous encounter or consult at a distance using ICTs.

In this paper we focus on the third type of telemedicine and we will use the term video consultations (VCs), which has been used extensively in the literature [2,4,5], to

[1] Corresponding Author: Maria Hägglund, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, 1st floor, 752 37 Uppsala, Sweden; maria.hagglund@kbh.uu.se.

distinguish these types of telemedicine solutions from e.g. chat functions [6], and telemonitoring [7].

In 2015, a new type of telemedicine service emerged in Sweden – patient-initiated, first-line VCs with primary care. Private online healthcare providers led the implementation, providing exclusively VCs, and were controversial. They were criticized on several points; draining the tax funded Swedish healthcare system by offering unnecessary care [8], poor quality due to difficulties assessing patients online, and unequal access to care. Limited research has however been performed regarding this novel form of care. Studies focusing on patients' experiences, indicate that patients' appreciate the easy access to care and are overall satisfied with the care they receive [9], similarly to other studies of VCs [10]. Elderly patients in Sweden have however expressed ambivalence towards VCs [11], confirmed by the annual "The Swedes and the Internet" survey which indicate that older age groups fall behind in the adoption of eHealth and VCs [12]. In a UK study, VCs were popular among some patients and staff, but also proved challenging to implement in "a busy and financially stretched acute hospital setting" [13]. In a Norwegian study of general practitioners' experiences of VCs during the Covid-19 pandemic, they expressed that VCs were more useful when they already knew the patient [4]. The context of online healthcare providers is however very different, and few studies have focused on healthcare professionals' experiences of this context. In this study, we interviewed physicians working at a Swedish private digital care provider (KRY AB). The aim of this study is to explore what improvements physicians' who have experiences of providing care through VCs for an online healthcare provider, would like for the future.

2. Methods

Qualitative interviews were performed with 22 physicians working for a Swedish online healthcare provider. The consolidated criteria for reporting qualitative research (COREQ) guideline [14] was used for reporting. The research team consist of 4 PhDs, all women, with experience of qualitative research. NF is also a medical doctor, and was at the time employed by the online healthcare provider from where study participants were included. AKE is a registered nurse, working both clinically and as a researcher.

The online healthcare provider delivers VCs with nurses, physicians, and psychologists available through a web-based and mobile platform in Sweden, making VCs possible via chat or video directly on a smartphone. Physicians have the choice to work from home or from the main office. All physicians working from home are provided with a pre-installed laptop to ensure centrally controlled security and updates. In addition to the VC applications, the physicians have access to a communication tool for collegial support and second opinions, the Swedish National Patient Overview (NPO) giving access to patients' electronic health records from across Sweden [15], and clinical guidelines for online care.

A convenience sample of volunteer participants were included in the study. Recruitment began in January 2019; information about the study was sent out by an administrator via email to all physicians at the healthcare provider, clearly stating that participation was voluntary. Thirty-three physicians expressed an interest; 11 later declined to do so for various reasons, lack of time being the most prominent. Table 1 gives an overview of the study participants' characteristics.

Even gender distribution was achieved with 11 female and 11 male participants. The majority of our study participants worked part-time (4-20 h/week) with VCs, and part time in traditional care. A few worked exclusively with VCs, part or full-time. Of these, a few had retired and now worked online from home. The majority worked in Sweden, but a few worked from abroad. Experience of working with VCs ranged between 5 months and 3 years (median 1 year).

Table 1. Overview of the 22 study participants' characteristics

Characteristic	Number (%)
Age	
30-39	5 (23%)
40-49	5 (23%)
50-59	4 (18%)
60-69	5 (23%)
70-79	3 (13%)
Level of medical training	
Specialist	15 (68%)
Resident	3 (14%)
Not specialist or resident	4 (18%)

A semi-structured interview guide was designed (MH, AKE and NF), tested and refined (MH, AKE and a health informatics master student). Here, we focus on the participants answers to our final questions regarding any improvements they would like to see in the future. Interviews were performed February-April 2019, before the pandemic increased the use of VCs dramatically. Each interview lasted between 30 minutes and 1 hour. Three researchers performed the interviews (MH, AKE and a health informatics masters student). The interviewer introduced themselves at the start of the interview, describing their interest in the topic at hand and experience of clinical work. Only the interviewer and the study participant were present during the interview. All interviews were performed online, recorded and later transcribed.

Data analysis was performed by all four co-authors (MH, AKE, ND and NF), following an inductive content analysis approach according to Graneheim and Lundman [16]. All authors read through all interviews, and then took responsibility for coding a subset of the interviews. Frequent meetings were held where codes were discussed and compared, and categories and themes emerged, until consensus was reached. Ethical approval for the study was granted by the Swedish Ethical Review Authority (Dnr: 2018/2318-31/5).

3. Results

Two themes related to future improvements emerged; blended care, and technical innovation.

In education, the term blended learning refers to education provided both online and on campus, to reap the benefits of both contexts [17]. The respondents in our study also expressed a need for further integration between digital and physical care, i.e. **blended care**.

"I think there needs to be a physical connection, so you do not only offer digital care to patients. It's better for patients if they can handle all their health issues, so if they are unable to get help online, that they can be referred to a physical appointment, but still within the same healthcare provider..." (Interview 22)

In addition to increased integration of online services within the private healthcare providers that dominated the online care market at the time of this study, participants also believed that in the future the regions, or public healthcare providers, would increase their offers of VCs, integrated with their regular services.

"Yes, and I actually think that this is part of the future of healthcare, in public healthcare, that such solutions will be integrated. The private online care providers have progressed fast and it will be integrated into public care safely." (Interview 22)

Of course, at the time writing this, we know that the Covid-19 pandemic has sped up this process, and VCs have now been implemented throughout Swedish public healthcare, both primary and secondary care.

In addition to the more organizational improvements of integrating physical and online care, our participants also expressed a need for **technical innovation**. Most participants expressed that digital care will be improved by ancillary technologies that can eliminate digital care limitations, and strengthen the entire digital care process.

"I hope for more analyzes that we can rely on. For example, a standard temperature gauge so we know at any time what the temperature is. There are also quick analyzes that can be linked to the mobile as well." (Interview 10)

4. Discussion

The physicians wanted to see more integration between online and traditional healthcare, to make telemedicine a normalized part of their everyday clinical work. Considering the development during the pandemic, this is now the case. In the USA, telemedicine – the use of phones and video visits – soared from fewer than 2% of primary care visits in 2019 to more than 35% by April 2020 [18]. In Sweden, the private online care providers increased their visits with 60% during the first months of the pandemic, but in parallel 19 (of 21) regions implemented their own telemedicine solutions which have been used extensively. In Region Stockholm, the use of the regions own solution for VCs (Alltid öppet) soared from 3000 visits in January 2020 to 36000 visits in April 2020 [19], and the adoption curve seem to have followed the increase of COVID cases [20].

In this study, we only interviewed physicians who worked fully (few individuals) or partly (the majority) for a private online healthcare provider. The needs for improvements they identified may be different from those of physicians who work with telemedicine or VCs as an integrated part of their work at a traditional primary care center. However, all of the participants had experience of working in traditional, physical and public health care and reflected on the differences between the contexts. Yet, further research is required into the differences between introducing VCs in a traditional healthcare organization and in an exclusively online organization.

5. Conclusion

The pandemic year 2020 has proven, nationally and internationally, to lead to a significant increase in the implementation and integration of VCs, in traditional care systems. Key aspects regarding the future of VCs as highlighted by our participants was the need for more collaboration between digital and physical care, that digital care should be offered by regions (public healthcare providers), and that technical improvements are needed to facilitate online assessments of patients.

Acknowledgement
AFA Insurance provided funding for the analysis (MH, ND) through the research project "ePrIm" (190210).

References

[1] Webster P. Virtual health care in the era of COVID-19, The Lancet, vol. 395, no. 10231, pp. 1180–1181, 2020, doi: https://doi.org/10.1016/S0140-6736(20)30818-7.
[2] Greenhalgh T, Wherton J, Shaw S, and Morrison C. Video consultations for covid-19, BMJ, vol. 368, 2020, doi: 10.1136/bmj.m998.
[3] Hashiguchi T C O. Bringing health care to the patient, OECD Health Working Papers, no. 116, 2020, doi: https://doi.org/10.1787/8e56ede7-en.
[4] Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, Getz L. Suitability of Video Consultations During the COVID-19 Pandemic Lockdown: Cross-sectional Survey Among Norwegian General Practitioners, J Med Internet Res, vol. 23, no. 2, p. e26433, 2021, doi: 10.2196/26433.
[5] James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, Scale-up, and Sustainability of Video Consulting in Health Care: Systematic Review and Synthesis Guided by the NASSS Framework., J Med Internet Res, vol. 23, no. 1, p. e23775, Jan. 2021,
[6] Cajander Å, Larusdottir M, Hedström G, The effects of automation of a patient-centric service in primary care on the work engagement and exhaustion of nurses, Quality and User Experience, vol. 5, no. 1, p. 9, 2020, doi: 10.1007/s41233-020-00038-x.
[7] Ding H, Chen SH, Edwards I, Jayasena R, Doecke J, Layland J, Yang IA, Maiorana A. Effects of Different Telemonitoring Strategies on Chronic Heart Failure Care: Systematic Review and Subgroup Meta-Analysis. J Med Internet Res. 2020 Nov 13;22(11):e20032. doi: 10.2196/20032.
[8] Ellegård LM, Kjellson G. Nätvårdsanvändare i Skåne kontaktade oftare vårdcentral, Läkartidningen, vol. 116, no. november, pp. 1–5, 2019. [In Swedish]
[9] Gabrielsson-Järhult F, Kjellström S, Josefsson KA. Telemedicine consultations with physicians in Swedish primary care: a mixed methods study of users' experiences and care patterns, Scandinavian Journal of Primary Health Care, vol. 39, no. 2, pp. 204–213, Apr. 2021
[10] Liddy C, Drosinis P, Keely E. Electronic consultation systems: worldwide prevalence and their impact on patient care—a systematic review, Family Practice, vol. 33, no. 3, pp. 274–285, Jun. 2016
[11] Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. "Having to learn this so late in our lives…" Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care, Scandinavian Journal of Primary Health Care, vol. 37, no. 1, pp. 41–52, Jan. 2019
[12] Internetstiftelsen, Svenskarna och internet 2019, 2019. [In Swedish]
[13] Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, Campbell-Richards D, Ramoutar S, Collard A, Hodkinson I. Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study. J Med Internet Res 2018;20(4):e150, doi: 10.2196/jmir.9897.
[14] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups, International Journal for Quality in Health Care, vol. 19, no. 6, pp. 349–357, Dec. 2007, doi: 10.1093/intqhc/mzm042.
[15] Inera AB, Nationell Patientöversikt (NPÖ). Accessed: Jan. 01, 2023. [Online]. Available: https://www.inera.se/tjanster/alla-tjanster-a-o/npo---nationell-patientoversikt/ [In Swedish]
[16] Graneheim UH, Lundman B, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, Nurse Educ Today, vol. 24, no. 2, pp. 105–112, Feb. 2004
[17] Bosch J, Ludwig C, Fluch-Niebuhr J, Stoevesandt D. Empowerment for the Digital Transformation: Results of a Structured Blended-Learning On-the-Job Training for Practicing Physicians in Germany, Int J Environ Res Public Health, vol. 19, no. 20, p. 12991, Oct. 2022
[18] Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and Content of Primary Care Office-Based vs Telemedicine Care Visits During the COVID-19 Pandemic in the US, JAMA Network Open, vol. 3, no. 10, pp. e2021476–e2021476, 2020
[19] Cederberg J. Sharp Increase in Digital Care (in Swedish, Kraftig ökning av digital vård)., *Läkartidningen*.
[20] Jäderlund Hagstedt L, Hägglund M, Petersson G, Davoody N, Hvitfeldt H. Impact of the Covid-19 pandemic on use of Video consultations in a Swedish Primary care setting, presented at the 18th Scandinavian Conference on Health Informatics, Aug. 2022, pp. 212–213. doi: 10.3384/ecp187037.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230315

U-Net-Based Segmentation of Current Imaging Biomarkers in OCT-Scans of Patients with Age Related Macular Degeneration

Kemal YILDIRIM[a,1], Sami AL-NAWAISEH [b], Sophia EHLERS [b], Lukas SCHIEßER[b], Michael STORCK[a], Tobias BRIX[a], Nicole ETER[b] and Julian VARGHESE[a]
[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department of Ophthalmology, University Hospital Münster, Münster, Germany*

Abstract. Age-related macular degeneration (AMD) is the leading cause of blindness in the Western world. In this work, the non-invasive imaging technique spectral domain optical coherence tomography (SD-OCT) is used to acquire retinal images, which are then analyzed using deep learning techniques. The authors trained a convolutional neural network (CNN) using 1300 SD-OCT scans annotated by trained experts for the presence of different biomarkers associated with AMD. The CNN was able to accurately segment these biomarkers and the performance was further enhanced through transfer learning with weights from a separate classifier, trained on a large external public OCT dataset to distinguish between different types of AMD. Our model is able to accurately detect and segment AMD biomarkers in OCT scans, which suggests that it could be useful for prioritizing patients and reducing ophthalmologists' workloads.

Keywords. Age-Related Macular Degeneration, Spectral Domain Optical Coherence Tomography, Convolutional neural Networks, Image Segmentation

1. Introduction

In western countries, age-related macular degeneration (AMD), is the primary cause of blindness in people over 65 affecting around 1 in 4 adults over the age of 75. A projected 196 million people will experience AMD by the year 2020, according to estimates. By 2040, this figure is anticipated to reach 288 million [1, 2].

International retinal specialists used fundus imaging to classify AMD based on clinical features to determine the likelihood that a late AMD will develop [3]. Depending on drusen and pigmentary changes within two disk diameters of the fovea, eyes were classified as normal age-related changes, early, moderate and late AMD. As new therapy modalities emerge aiming to prevent the development of a late AMD, it crucial to precisely detect and follow up the changes in drusen and pigmentary changes, to predict the natural disease progression and thereby assess the therapeutic benefit.

The quantitative morphological evaluation of drusen and pigmentary changes is made possible by several established and emerging imaging technologies, including

[1] Corresponding Author: Kemal Yildirim, E-mail: kemal.yildirim@uni-muenster.de.

color fundus photography (CFP), fundus autofluorescence (FAF), infrared imaging (IR) and spectral domain optical coherence tomography (SD-OCT). SD-OCT imaging offers the most precise and prompt diagnosis. Compared to FAF and IR, which may show higher variability since the image intensities substantially fluctuate due to variations in illumination and corneal curvature, SD-OCT scans offer consistent anatomic landmarks for objective assessments.

OCT is a non-invasive imaging method that uses low-coherence light to produce cross-sectional images of the macula or optic nerve head [4]. OCT is a highly favored method by ophthalmologists for the evaluation of retinal diseases, such as AMD, due to its non-invasiveness and simplicity of image acquisition [5]. However, it takes a lot of time and effort for ophthalmologists to precisely examine several OCT cross-sections for each patient. Additionally, the chronic nature of AMD adds to the load on ophthalmologists and medical facilities. Therefore, the availability of a computer-aided diagnosis (CAD)-based screening tool that is automated could aid in prioritizing patients based on their conditions and lessen this load.

2. Methods

The open-source Computer Vision Annotation Tool (CVAT) [6] was used by a trained domain expert to label four types of biomarkers on 1200 OCT scans: drusen with 4181 labels, pseudodrusen with 108 labels, choroidal neovascularizations (CNV) with 2810 labels, and pigment epithelial detachments (PED) with 86 labels. To ensure proper capture of the structures of interest, the annotations were reviewed and the polygons were refined by two additional clinicians, if necessary. This was especially helpful in wet AMD with severe geographic atrophy, when the retinal layers were no longer easy to distinguish from each other. Nevertheless, to obtain a sample of clinical cases that is as representative as possible, cases with significant fibrotic or atrophic lesions were also included. Additionally, we have used a range of data augmentation techniques to address the extreme label imbalance and improve its ability to generalize to new data. Following transformations were applied to a copy of the original dataset, prior to the training process, ensuring that the input for all training iterations remained consistent: Gaussian noise, contrast adjustments, elastic deformations, grid/optical distortions and random affine transformations [7].

Our segmentation model is built on CNN with U-net and was trained on the data mentioned above. Different studies have demonstrated that U-net with an additional attention module and ResNet34 as encoder is an effective combination to handle datasets with imbalanced classes and minimize loss [8–10]. For comparison purposes, the encoder was initialized with ImageNet and custom weights from a public OCT-Dataset, which contains 108309 OCT scans [11]. This usually reduces the amount of data necessary for training, allowing models to learn task-specific features while utilizing the knowledge gained from broader datasets.

To calculate the weights of the custom dataset, a classifier based on the same ResNet34 encoder of our segmentation model was trained to distinguish between the given classes of diabetic macular edema (DME), drusen (dry AMD), CNV (wet AMD) and normal (healthy). Our ResNet34-Classifier was pretrained on ImageNet weights. A weighted cross-entropy loss with inverse class frequencies was used to adjust the relative contribution of each class to the loss computation. The weights of the classifier's encoder were then extracted and transferred to our segmentation model’s encoder.

In addition, our segmentation model used an alpha-balanced compound loss with Dice and weighted Focal loss to train the network to successfully segment small pathogenic structures within the OCT scans [12]. Specific combinations of loss functions have shown to improve focus on relevant regions and make more precise predictions.

To measure the performance of our model, the Dice/F1 score was calculated for each class and epoch, as the structures for CAD must be captured not only completely but also precisely. This metric measures model accuracy by combining its precision and recall, indicating overlap between predicted segmentation and ground truth mask.

3. Results

Classifier (OCT). The classifier is built on top of ResNet34 and was initialized using ImageNet weights. It was trained over 15 epochs and by using 16 training examples per mini-batch from the public dataset. Initial learning rate was 1e-4. At epoch 7, the weighted cross-entropy loss was at its minimum. With an 80:20 split for each class, the training accuracy reached 98.207% and the validation accuracy peaked at 97.697%. Rounded accuracies for each class were as follows: NORMAL 98.49%, DME 97.9%, DRUSEN 96.54%, CNV 95.23%.

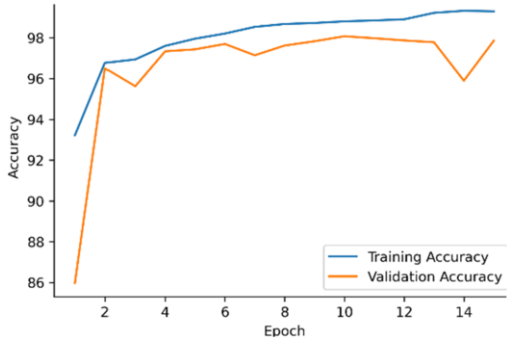


Figure 1. Training and validation accuracy per epoch

Segmentation Model. We have tested our segmentation model with three different weight initialization strategies: random initialization, pre-trained ImageNet weights and custom weights from our OCT classifier. Table 1 compares the results from the epochs with the highest F1 scores. Figure 2 exemplifies results on a separate test dataset.

We used a mini-batch size of 4, an initial learning rate of 1e-3. and 40 epochs. The same preprocessing and parameters were used throughout all three trials. The compound loss was defined and balanced as described in Eq. (1).

$$Loss_{Compound} = \alpha \cdot Loss_{Focal} + (1-\alpha) \cdot Loss_{Dice}\,, \quad \alpha = 0.6, \quad \gamma_{Focal} = 2 \qquad (1)$$

Table 1. The F1 scores of the best epoch in each case for all three executions

	Epoch	Drusen	CNV	PED
Random Initialization	17/40	0.777	0.752	0.899
ImageNet	13/40	0.763	0.832	0.888
OCT	21/40	0.807	0.844	0.878

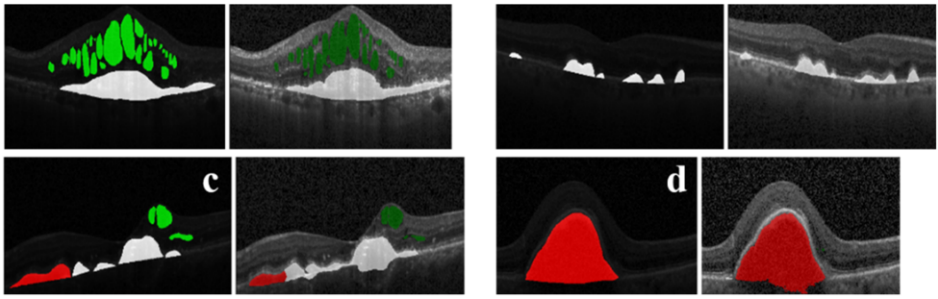


Figure 2. Ground truth and prediciton with pretrained encoder. Drusen (white), CNV (green) and PED (red).

4. Discussion

This study's main objective is to analyze the feasibility of CAD and progress monitoring for patients with AMD. We therefore utilize a broader set of features than previous OCT segmentation research, which mostly focuses on single biomarkers such as fluid, layers, or drusen [13]. In addition, we investigate the impact of using weights from different datasets on the segmentation of relevant biomarkers collectively.

Given several AI-based systems achieving comparable image discrimination rates as that of retinal specialists [14], we primarily focus on improving the segmentation. The underlying hypothesis was that it may be improved by using transfer learning with weights from similar or unrelated datasets. Thus, three image segmentation models were trained and compared for the classes drusen, CNV, PED and pseudodrusen. Pseudodrusen had to be excluded due to the insufficient number of labels and the resulting bias in the loss computation.

As seen in the Table 1, the model trained using OCT-weights had the highest F1 score of 0.807 for drusen class. This suggests that training on the custom dataset may have improved the model's class segmentation. The comparably low value of drusenoid structures is likely since the pigments within these deposits are similar to those found in the choroid beneath. This makes it difficult to accurately label and segment lower parts of drusen, as it is based on the assumption of how the retinal pigment epithelium was prior to the formation of a druse. This becomes especially challenging in cases of wet AMD with large and overlapping deposits.

The model trained with our custom weights had the highest F1 score for CNV, 0.844. This suggests that our encoder provided adequate training for this class. ImageNet alone resulted in the highest F1 score for PED, 0.888. OCT-weighted model placed second with 0.878. Transfer learning may have improved performance, but our encoder was unable to produce any additional improvement. The higher F1 score seen in CNV and PED could be related to the hypopigmentation of these structures, which allows a better differentiation from the surrounding tissue. The background class was disregarded from Table 1 due to many easy positives resulting in a F1 score near to 1.

The precision of segmentation results is often influenced by the characteristics of the datasets used. We have included a substantial number of wet AMD cases (600 out of 1300), which may have a diminishing but important impact on the research results.

It is also essential to note that no systematic hyperparameter tuning was performed. Instead, previously established parameters from similar cases were used for our classifier

and segmentation model, including those for loss computation. Our initial tests have demonstrated that the compound and balance of loss functions used was sufficient but with obvious room for enhancement.

5. Conclusion and Outlook

Using image segmentation models, this study evaluated the viability of CAD for patients with AMD. The results demonstrated that transfer learning with weights from similar or unrelated datasets can improve the performance of models for specific classes. Since our classifier has been trained on a publicly available dataset, it can therefore be tested on various diagnoses and biomarkers in future research. In our case, additional data labelling for existing and new AMD biomarkers, preprocessing, and augmentation, as well as hyperparameter tuning, may improve the results. We intend to create interpolated 3D volumetric scans of the OCT images to extract additional properties such as volume, diameter and other criteria that are necessary for clinical classification of AMD [3].

References

[1] Friedman DS, O'Colmain BJ, Muñoz B, Tomany SC, McCarty C, DeJong, Paulus T. V. M., et al. Prevalence of Age-Related Macular Degeneration in the United States. Archives of Ophthalmology. 2004;122:564–72. doi:10.1001/archopht.122.4.564.

[2] Laurence S Lim, Paul Mitchell, Johanna M Seddon, Frank G Holz, Tien Y Wong. Age-related macular degeneration. The Lancet. 2012;379:1728–38. doi:10.1016/S0140-6736(12)60282-7.

[3] Ferris FL3, Wilkinson CP, Bird A, Chakravarthy U, Chew E, Csaky K, Sadda SR. Clinical classification of age-related macular degeneration. Ophthalmology. 2013;120:844–51. doi:10.1016/j.ophtha.2012.10.036.

[4] van Velthoven, Mirjam E J, Faber DJ, Verbraak FD, van Leeuwen TG, Smet MD de. Recent developments in optical coherence tomography for imaging the retina. Prog Retin Eye Res. 2007;26:57–77. doi:10.1016/j.preteyeres.2006.10.002.

[5] Vineeta Das, Samarendra Dandapat, Prabin Kumar Bora. Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images. Biomedical Signal Processing and Control. 2019;54:101605. doi:10.1016/j.bspc.2019.101605.

[6] CVAT.ai Corporation. Computer Vision Annotation Tool (CVAT); 2022.

[7] Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: Fast and Flexible Image Augmentations. Information 2020. doi:10.3390/info11020125.

[8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015: arXiv. doi:10.48550/ARXIV.1505.04597.

[9] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition 2015: arXiv. doi:10.48550/ARXIV.1512.03385.

[10] Roy AG, Navab N, Wachinger C. Recalibrating Fully Convolutional Networks with Spatial and Channel 'Squeeze & Excitation' Blocks 2018: arXiv. doi:10.48550/ARXIV.1808.08127.

[11] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018;172:1122-1131.e9. doi:10.1016/j.cell.2018.02.010.

[12] Yeung M, Sala E, Schönlieb C-B, Rundo L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation 2021: arXiv. doi:10.48550/ARXIV.2102.04525.

[13] Liefers B, Taylor P, Alsaedi A, Bailey C, Balaskas K, Dhingra N, et al. Quantification of key retinal features in early and late age-related macular degeneration using deep learning. American Journal of Ophthalmology. 2021;226:1–12.

[14] Sheng B, Chen X, Li T, Ma T, Yang Y, Bi L, Zhang X. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. Frontiers in Public Health. 2022;10.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230316

Random Forest and Gradient Boosted Trees for Patient Individualized Contrast Agent Dose Reduction in CT Angiography

René PALLENBERG[a,b,1], Marja FLEITMANN[c], Andreas Martin STROTH[d], Jan GERLACH[d], Alexander FÜRSCHKE[d], Jörg BARKHAUSEN[d], Arpad BISCHOF[d,e] and Heinz HANDELS[b,c]
[a] *University of Lübeck, Institute of Signal Processing, Germany*
[b] *University of Lübeck, Institute of Medical Informatics, Germany*
[c] *German Research Center for Artificial Intelligence, Artificial Intelligence in Medical Imaging, Lübeck, Germany*
[d] *University Medical Center Lübeck, Department of Radiology and Nuclear Medicine*
[e] *IMAGE Information Systems Europe, Rostock, Germany*

Abstract. This work aims to recognize the patient individual possibility of contrast dose reduction in CT angiography. This system should help to identify whether the dose of contrast agent in CT angiography can be reduced to avoid side effects. In a clinical study, 263 CT angiographies were performed and, in addition, 21 clinical parameters were recorded for each patient before contrast agent administration. The resulting images were labeled according to their contrast quality. It is assumed that the contrast dose could be reduced for CT angiography images with excessive contrast. These data was used to develop a model for predicting excessive contrast based on the clinical parameters using logistic regression, random forest, and gradient boosted trees. In addition, the minimization of clinical parameters required was investigated to reduce the overall effort. Therefore, models were tested with all subsets of clinical parameters and each parameter's importance was examined. In predicting excessive contrast in CT angiography images covering the aortic region, a maximum accuracy of 0.84 was achieved by a random forest with 11 clinical parameters; for the leg-pelvis region data, an accuracy of 0.87 was achieved by a random forest with 7 parameters; and for the entire data set, an accuracy of 0.74 was achieved by gradient boosted trees with 9 parameters.

Keywords. Random Forest, Gradient Boosted Trees, Logistic Regression, CT angiography, Contrast

[1] Corresponding Author: René Pallenberg, E-mail: r.pallenberg@uni-luebeck.de.

1. Introduction

Computed tomography angiography (CTA) is an important method for detecting diseases such as tumors, pulmonary embolism and vascular stenosis [1]. Administration of contrast agents is necessary for CTA. Many of these contrast agents contain iodine, which may harm the patient's body. Observed side effects are allergic reactions, hyperthyroidism, and deterioration of renal function up to renal insufficiency [2].

One way to avoid side effects is to reduce the administered dose of contrast agents [3]. Currently, standard doses are usually administered in everyday clinical practice. The dose is often higher than required for sufficient contrast quality. This is to avoid generating an insufficiently contrasted image, which would result in repetition and thus in a renewed exposure of the patient.

In this paper, we present methods to recommend contrast dose reduction for patients on an individual basis. To do this, the system is trained to predict whether the contrast quality of a CTA image is "excessive" based on clinical parameters recorded before contrast agent administration [4]. It is assumed that if a CTA image had excessive contrast, the contrast dose could have been reduced [5]. It is important to achieve high precision to ensure that no CTA images are created with insufficient contrast. In addition, the influence of each clinical parameter in the prediction models was investigated.

2. Material and Methods

2.1. Data

A clinical study was conducted by the Department of Radiology and Nuclear Medicine at the University Hospital Schleswig-Holstein in Lübeck, Germany [4]. In this study, additional clinical parameters were recorded from patients who had a CTA performed as part of their treatment. They were recorded before the administration of the contrast agent.

These parameters were: ankle-brachial index (ABI), age (AGE), body mass index (BMI), blood pressures diastolic (BPD) and systolic at rest (BPS), 5 min (BPD5, BPS5) and directly (BPD0, BPS0) before administration, creatine (C), gender (G), glomerular filtration rate (GFR), γ-glutamyltransferase (GGT), height (H), hematocrit (HC), hemoglobin concentration (HB), oxygen saturation (OS), Pulse directly (P0) and 5 min (P5) before administration, weight (W) and waist circumference (WS). In total, the data set comprises 263 CTA images and the corresponding clinical parameters. The images were labeled as "excessive" based on the quality of their contrast [5-7]. Excessive means that the contrast is higher than necessary. To determine this, the mean HU values of selected regions of interest were determined by a radiologist [6]. The data set includes CTA images of eight different recording protocols covering different body regions.

Table 1. Distribution for the used data sets.

Data set	Not Excessive	Excessive	Total
Aorta (AORTA)	22	68	90
Leg and pelvic vessel (LPV)	48	22	70
All protocols combined (COMB)	107	156	263

Three data sets were used for the experiments, one containing the aortic protocol (AORTA), one the leg and pelvic vessel protocol (LPV), and one all eight protocols combined (COMB), see Table 1.

2.2. Classification

To predict whether the contrast of a CTA image is "excessive" or not, three classification methods widely used in tabular data in medicine and biology were applied [8,9]: Random Forest, Gradient Boosted Trees, and Logistic Regression as established reference method [10-12]. The 21 clinical parameters of the patient belonging to the CTA served as input for the classifiers. Training and evaluation were performed using leave-one-out crossvalidation, for each of the three methods. A grid search was performed to optimize the hyperparameters of the classifiers. Accuracy was chosen as the optimization criterion. A preselection of input parameters was investigated to further improve the prediction and reduce the number of required clinical parameters. Therefore, the methods were tested for all possible subsets of the 21 clinical parameters. In addition, the importance of each clinical parameter for prediction was assessed by permutation and Gini importance [13]. The importance was predicted separately for the 50 models that achieved the highest accuracy on the respective dataset. For the overall importance, the results of these 50 models were averaged.

3. Results and Discussion

3.1. Classification Performance

The described evaluation procedure was done separately for the two subsets AORTA and LPV and also for all available data (COMB). Due to the unbalanced distribution of classes in the data sets, the Matthews correlation coefficient (MCC) was used as a measure of model performance. In Figure 1 the highest achieved MCC is related to the associated values for precision and the number of clinical parameters used.

This work should improve previous prediction methods of contrast quality. In [4] we show that a k-nearest neighbor method achieved an MCC of 0.34, an accuracy of 0.78 and a precision of 0.85 on the AORTA data set. By applying random forest, gradient boosted trees, and logistic regression and reducing the clinical parameters used, higher accuracy was achieved for the AORTA data set. The highest accuracy (0.84), MCC (0.54), and precision (0.86) were achieved by random forest using 11 clinical parameters (ABI, AGE, G, GFR, GGT, H, HB, P0, P5, W, WS). With the LPV data set, both gradient boosted trees with 7 (ABI, BMI, BPS, G, OS, P5, W) and random forest with 7 (ABI, AGE, BPD5, BPS5, GFR, HB, WS) clinical parameters achieved an MCC of 0.70 with a precision of 0.93 and an accuracy of 0.87. On the whole data set (COMB), the maximum MCC was 0.46 with a precision of 0.75 and an accuracy of 0.75, achieved by gradient boosted trees and 9 (ABI, AGE, BMI, BPD, BPD5, GFR, GGT, HC, OS, W) clinical parameters.

This shows that logistic regression cannot compete with the other two methods. On all data sets, the best results were achieved with 5-11 parameters. In addition to the reduced effort, the reduction of the 21 parameters also offers quality improvements. The differences in results between the data set LPV and AORTA can be explained by the distribution of the data. It is also seen that the results for COMB are much less.

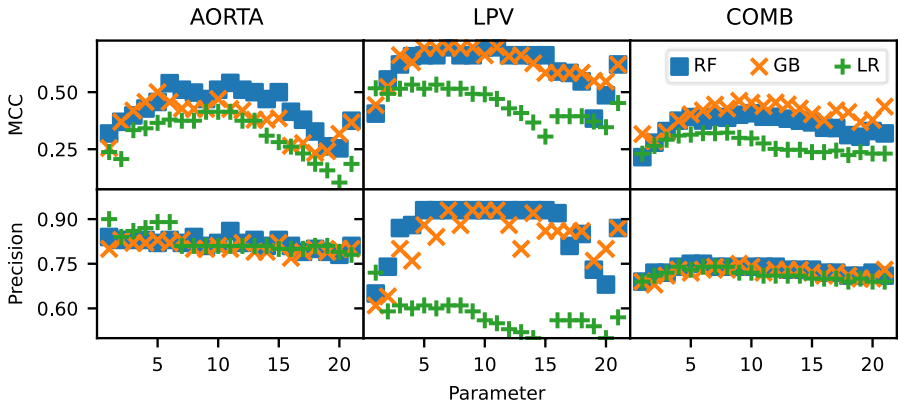


Figure 1. MCC and precision of the best model by MCC-Score, divided by the method (random forest (RF), gradient boosted trees (GB), and logistic regression (LR)), the data set containing aortic (AORTA), leg and pelvic vessel (LPV), and all eight protocols(COMB), and the number of clinical parameters used (1-21).

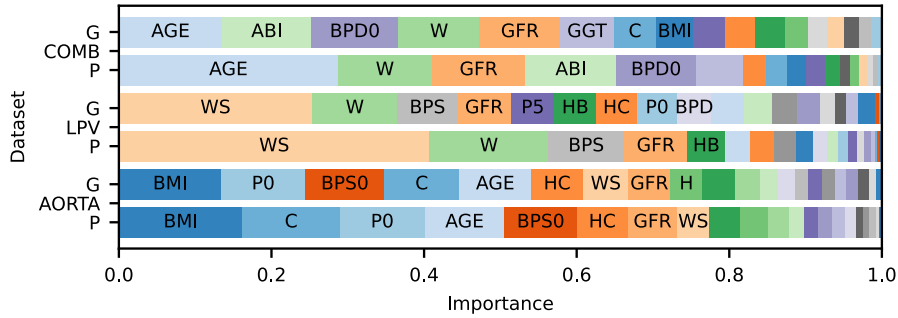


Figure 2. Resulting scaled importance of clinical parameters predicted by permutation (P) and Gini (G) importance.

3.2. Importance of the Clinical Parameters

To further investigate the relationship between the clinical parameters and the quality of contrast in the CTA images permutation and Gini importance was computed [13]. When looking at Figure 2, it can be seen that between 5 and 9 clinical parameters are rated as important and the others as hardly important. The results of the two methods differ only slightly in the evaluation of the importance of the parameters, which indicates the reliability of the results.

Depending on the data set, however, the important parameters differ. Also, the distance between the important and unimportant parameters is significantly greater in LPV and COMP than in AORTA. Therefore, generalization over different recording protocols is not feasible. On AORTA, AGE, BMI, BPS0, C, G, GFR, HC, P0, and WS were the most important parameters. On LPV, WS was by far the most important parameter W, BP5, and GFR were the next most important. On COMP, AGE, ABI, GFR, PD0 and W were the most important parameters.

4. Conclusion

This work aimed to develop improved recommendation models for contrast agent dose reduction. This is done by predicting excessive contrast of CTA images based on 21 clinical parameters of a patient. To build a predictive model three different methods gradient boosted trees, random forest, and logistic regression were tested. These methods were tested on three different data sets, AORTA, LPV, and COMB. Overall, a maximum MCC of 0.54 was achieved on the AORTA data set, with a precision of 0.86 and an accuracy of 0.84. This was achieved by a random forest using 11 clinical parameters. This represents an improvement to the results of our previous work, where a maximum MCC of 0.34 was achieved [4].

According to the experts, an accuracy of more than 90% and a precision of 95% should be achieved for clinical application. Even though this work has already improved the results, they still need to be increased. Our experiments show that generalization across multiple imaging protocols leads to poorer predictions. Therefore, the models should be built for each protocol and body region. Feature selection has shown that using about 5 to 11 clinical parameters as input leads to as good or better results than using all 21 clinical parameters. The results obtained may be used to reduce the number of clinical parameters to be included in the future.

References

[1] Yang X, Tridandapani S, Beitler JJ, Yu DS, Yoshida EJ, Curran WJ, et al. Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. Medical Physics. 2012;39(9):5732-9.

[2] Andreucci M, Solomon R, Tasanarong A. Side Effects of Radiographic Contrast Media: Pathogenesis, Risk Factors, and Prevention. BioMed Research International. 2014.

[3] Roos JE, Desbiolles LM, Weishaupt D, Wildermuth S, Hilfiker PR, Marincek B, et al. Multi-detector Row CT: Effect of Iodine Dose Reduction on Hepatic and Vascular Enhancement. RöFo – Fortschritte auf dem Gebiet der R¨ontgenstrahlen und der bildgebenden Verfahren. 2004;176(4):556-63.

[4] Fleitmann M, Soika K, Stroth A, Gerlach J, Fürschke A, Hunold P, et al. Computer-Assisted Quality Assessment of Aortic CT Angiographies for Patient-Individual Dose Adjustment. Studies in health technology and informatics. 2020;270:123-7.

[5] Becker CR, Hong C, Knez A, Leber A, Bruening R, Schoepf UJ, et al. Optimal Contrast Application for Cardiac 4-Detector-Row Computed Tomography. Investigative Radiology. 2003;38(11):690-4.

[6] Pallenberg R, Fleitmann M, Soika K, Stroth AM, Gerlach J, Fürschke A, et al. Automatic quality measurement of aortic contrast-enhanced CT angiographies for patient-specific dose optimization. International Journal of Computer Assisted Radiology and Surgery. 2020;15(10):1611-7.

[7] Panetta K, Gao C, Agaian S. No reference color image contrast and quality measures. IEEE Transactions on Consumer Electronics. 2013;59(3):643-51.

[8] Patel SJ, Chamberlain DB, Chamberlain JM. A Machine Learning Approach to Predicting Need for Hospitalization for Pediatric Asthma Exacerbation at the Time of Emergency Department Triage. Academic Emergency Medicine. 2018;25(12):1463-70.

[9] Parreco J, Soe-Lin H, Parks JJ, Byerly S, Chatoor M, Buicko JL, et al. Comparing Machine Learning Algorithms for Predicting Acute Kidney Injury. The American Surgeon. 2019;85(7):725-9.

[10] Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.

[11] Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. 2000;29.

[12] Hoffman JIE. Logistic Regression. In: Biostatistics for Medical and Biomedical Practitioners. Elsevier; 2015. p. 601-11.

[13] Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics. 2018 Nov;34(21):3711-8.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230317

V-care: An Application to Support Lifestyle Improvement in Children with Obesity

Cristiana LARIZZA[a,1], Silvana QUAGLINI[a], Michelangelo CHASSEUR[b], Valentina Bevolo[b], Gianvincenzo ZUCCOTTI[c,d] and Valeria CALCATERRA[c,e]
[a]*Dept of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy*
[b]*Kotuko srl, Italy*
[c]*Pediatric Department, "Vittore Buzzi" Children's Hospital, Milano, Italy*
[d]*Department of Biomedical and Clinical Science, University of Milano, Italy*
[e] *Department of Internal Medicine and Therapeutics, University of Pavia, Italy*

ORCiD ID: Cristiana Larizza https://orcid.org/0000-0002-5055-8665, Silvana Quaglini https://orcid.org/0000-0002-4288-2561, Michelangelo Chasseur https://orcid.org/0000-0002-9246-0746, Gianvincenzo Zuccotti https://orcid.org/ 0000-0002-2795-9874 , Valeria Calcaterra https://orcid.org/ 0000-0002-2137-5974

Abstract. Obesity is increasing in the pediatric population and it represents an important risk factor for the life-long development of several diseases. The aim of this work is to reduce children obesity through an educational program delivered through a mobile application. Novelties of our approach are the involvement of the families in the program and a design inspired to psychological/behavioral change theories, with the aim of maximizing the chance of patients' compliance to the program. A pilot usability and acceptability study has been performed on ten children aged 6-12 years using a questionnaire to evaluate eight system features on a Likert scale from 1 to 5. Encouraging results were obtained: mean scores were all above 3.

Keywords. Mobile application, Mobile health, Chatbot, Pediatric obesity

1. Introduction

Overweight and obesity impair health, leading to increased morbidity, premature mortality and economic burden for health and social system. Over the past 40 years, prevalence of overweight and obesity increased in many countries, in both children and adults [1]. Childhood obesity is estimated to increase by 60% over the next decade, reaching 250 million by 2030. This dramatic trend brings out the urgency for health promotion from childhood, before a pathology sets in.
Lifestyle modification, including diet and exercise, continues to be the mainstay of obesity prevention and treatment; unfortunately, lifestyle modification programs are often unsuccessful [2]. Reasons could be: lack of results in a short time, bad family habits, lack of motivation, low frequency of check-ups (often only every six months or more) that put the treatment burden mainly on the family and on the child himself.

[1] Corresponding Author: Cristiana Larizza, E-mail: cristiana.larizza@unipv.it.

All these findings suggested the development of a mHealth platform supporting both children and their families in following a treatment program and help physicians in monitoring goals achievements. As reported in the literature, mHealth has the potential for improving patient management when compared with standard care [3]. However, data on the usefulness of mHealth in pediatric obesity management are limited [4-7].

This paper presents V-care, an app purposely developed to offer effective interaction with young subjects, and the results of a pilot study conducted in 10 children.

2. Requirement Analysis and Design

2.1. Target population

When developing an application for young patients, it's crucial to decide the target age range. In fact, although pediatric age is broadly defined as 0-18 years, it is clear that, in that range, cognitive abilities change a lot over years, so that technological solutions can address only parents (for very young children), only children (for adolescents) or both. Since one-fits-all solution does not work, we decided to develop an app for children aged 6-12 years, i.e. very young but able to understand some basic lifestyle principles. This age requires two precautions: using a captivating and intuitive language appropriated for very young patients and involving their parents. The latter is mandatory because of two reasons, first family-based behavioral treatment is an evidence-based intervention for childhood obesity [8] and second the children may not have a personal mobile device yet, and/or could be unable to interact appropriately (reading/writing) with the app.

2.2. Basic requirements of the application

The review of the above-cited literature on systems designed to support young people in overcoming their state of obesity has guided us in defining the basic requirements to be included in V-Care. They are summarized in Table 1 and elaborated below.

Table 1. Basic requirements

1. Provide physicians with a monitoring dashboard	5. Monitoring both behavior and engagement
2. Involve the entire family	6. Tailoring the app to specific situations
3. Provide education on healthy lifestyle	7. Promote motivation and avoid drop-outs
4. Provide behavioral goals to the user	8. Allow different modalities of interaction

1. Physicians will see the data through their dashboard, in particular they will monitor BMI, the objective measure used to evaluate the children's compliance to the program. Periodical update of height and weight will be performed by parents.
2. Given the age range 6-12yrs, parents play an indispensable role in the success of the treatment. The system requires them to register the children, provide an initial profile for them, and regularly visualize the trends of the data managed by the app. This is the way to involve the family in their children's journey against obesity.
3. The app includes an educational section that, besides reporting certified information about the best dietary behaviors, is made more appealing by a set of success stories. A story can also be the experience of an app users himself, and this can both motivate the children to improve their health status and the future users to engage.

4. The app is compliant with the latest guidelines on children obesity, providing general goals that every patient should achieve regarding nutrition, physical activity and sleep, e.g., they suggest the correct number of portions for vegetables, glasses of water, fruit.
5. Since face-to-face visits are only one or twice a year, for a tighter monitoring the app administers monthly questionnaires providing validated behavioral scores, and also provides a light diary to collect information about the child's habits daily or weekly.
6. The app is conceived as a virtual coach that reacts appropriately according to the patient's performance by providing some feedback. Good child performance is rewarded, while failures trigger encouragement to the child to do better in the future.
7. To face drop-out for loss of motivation, we must strengthen subjects' engagement with effective strategies. First, at registration the child may choose his preferred avatar (a pet). Moreover, the app implements gamification, e.g., a quiz section that gives medals or rewards if the child provides correct answers. This strategy allows to enjoy learning the fundamentals of healthy eating and the importance of physical activity.
8. To make the app usage more friendly, the innovative technique of conversational agents (chatbot) has been exploited. It is used to collect responses to questionnaires and quizzes, to give specific "practical advice" about healthy cooking, or correct behavior. The language and the kind of interaction of the chatbot with children is designed to mimic a mate who addresses them in a friendly tone and tries to encourage them.

2.3. Functional and technical Architecture

Figure 1 shows the architecture of V-Care, which can be used on IOS and Android operative systems. Three user categories exist: children/parents who will use a smartphone, doctors and administrators who access the web application from a PC, each of them accessing different features. Parents register their children in the app and provide short demographics, height, weight and one or more referral doctors. Furthermore, each child is associated with his parent's account, but he could use his personal smartphone, if he has one, and choose his favorite avatar. Subjects are then asked to periodically fill in questionnaires for lifestyle monitoring. All features in section 2.2 are implemented.

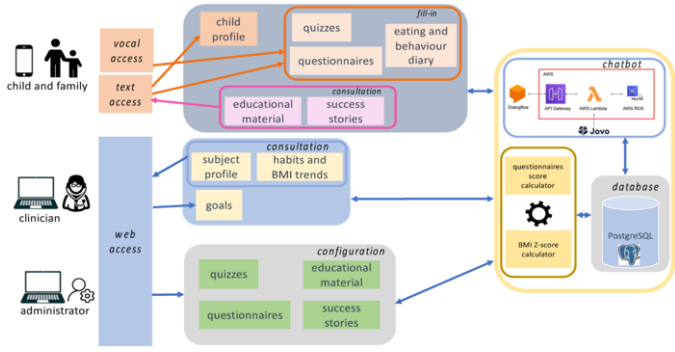


Figure 1. The App architecture. Technical details in the right: the chatbot is implemented using Dialogflow and AWS services: API Gateway to route the chatbot request to the correct Lambda function (written in Jovo) that will prepare the answer to the user. The Lambda function accesses the RDS (database) to get the necessary information and fulfill the request. The same mechanism is used both to manage the text and the voice based interaction.

The chatbot module provides new means for interacting with some parts of the application, through vocal commands or keyboard. The bot acts a child's friend, maintaining an informal and engaging language, and entertaining the user, even while completing a questionnaire; e.g., it encourages the child to continue filling in by indicating how many questions are missing, or if he reached halfway through the questionnaire, and at the end of the compilation, the score is calculated and a feedback is given based on the comparison with previous scores (if any). As for the "practical advice" part, different conversation flows between the chatbot and the user have been developed. The user can ask the chatbot for advice on some topics that will be presented by the chatbot itself. Those concern nutrition, physical activity and cooking, which expand into sports, a sedentary lifestyle, healthy cooking methods, "special recipes" and foods that the user does not like. The idea is to give simple explanations and provide practical and qualitative advices. Moreover, intolerances and disliked foods and sports are stored in the DB in order to personalize the conversation.

Doctors have access to a dashboard showing the BMI trend of the subjects, the data entered daily or weekly in the personal diary and the questionnaire answers and scores.

Administrators can manage: the registration of the doctors and of the admin in the back office, the questionnaires and quizzes to be proposed to users, with the related scores, educational material and curiosities, success stories and the guidelines to follow in terms of nutrition and behavior.

3. The Developed Application and the Validation Study

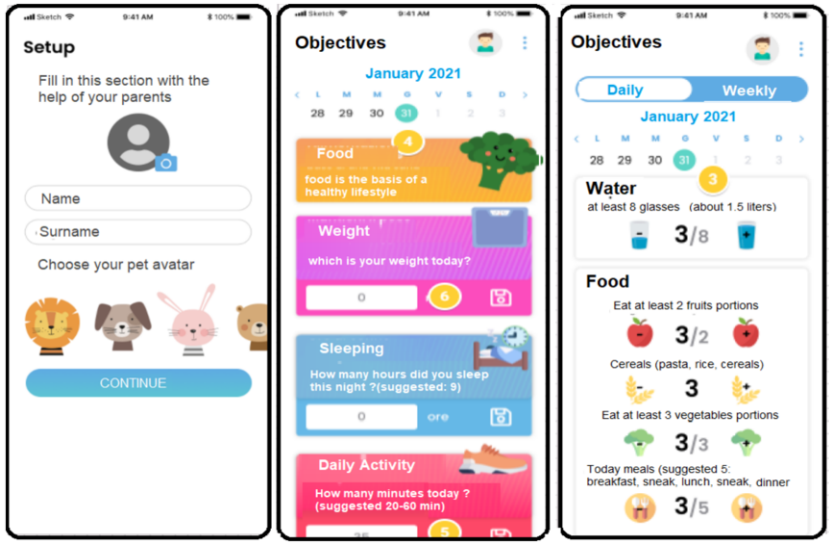


Figure 2. A Three screenshots of the children's and their parents' GUI.

Figure 2 shows the graphical interface of the app used by children and their parents. Ten children (6M/4F, 9.5±1.3 yrs) with obesity (BMI ε2) have been enrolled in the usability study. The parents gave consent to participate (protocol number 2020/ST/298, approval date 01/12/20). After an initial training by the doctors, they utilized the app at home for a period of two weeks, at the end of which they were required to fill in an anonymous

evaluation questionnaire. For each question they had to provide a score from 1 to 5, where 1 is the worst rating. Figure 3 shows the results.
The app resulted in a very good overall friendliness and perceived usefulness, even if some criticalities emerged particularly for the curiosity section and the chatbot.

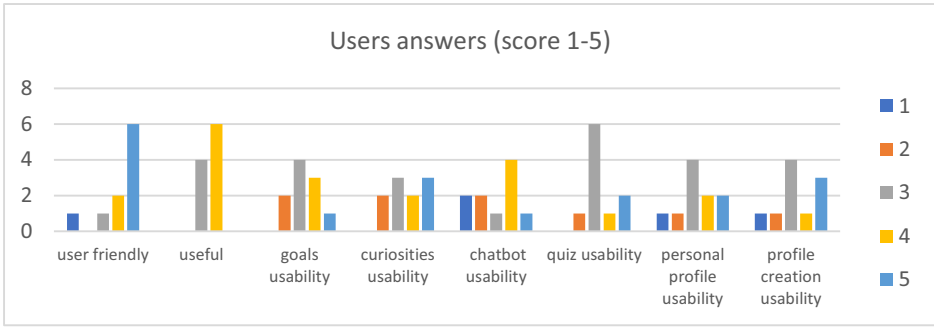


Figure 3. Results of the usability evaluation study: number of patients assigning the different scores for each feature

4. Conclusion and Future Work

V-care has been created by a multidisciplinary team with expertise in pediatric obesity and home-monitoring apps, and offers a dedicated and easy tool to improve pediatric patients management. A usability study provided encouraging results. Limitations of the study are the small sample size, the short time evaluation and the limited exploitation of the vocal interaction, until now performed only at experimental level. Validated instruments will be used in further evaluation studies.

References

[1] https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight
[2] Valerio G, Maffeis C, Saggese G, et al. Diagnosis, treatment and prevention of pediatric obesity: consensus position statement of the Italian Society for Pediatric Endocrinology and Diabetology and the Italian Society of Pediatrics. Ital J Pediatr. 2018;44(1):88.
[3] Arthurs N, Tully L, O'Malley G, Browne S. Usability and Engagement Testing of mHealth Apps in Paediatric Obesity: A Narrative Review of Current Literature. Int J Environ Res Public Health. 2022;19(3):1453.
[4] Jelalian E, Darling K, Foster GD, Runyan T, Cardel MI. Effectiveness of a Scalable mHealth Intervention for Children with Overweight and Obesity. Child Obes. 2022;
[5] Reis CI, Pernencar C, Carvalho M, et al. Development of an mHealth Platform for Adolescent Obesity Prevention: User-Centered Design Approach. Int J Environ Res Public Health. 2022;19(19):12568. Published 2022 Oct 1.
[6] de Arriba Muñoz A, García Castellanos MT, Cajal MD, Beisti Ortego A, Ruiz IM, Labarta Aizpún JI. Automated growth monitoring app (GROWIN): a mobile Health (mHealth) tool to improve the diagnosis and early management of growth and nutritional disorders in childhood. J Am Med Inform Assoc. 2022;29(9):1508-1517.
[7] De-Jongh González O, Tugault-Lafleur CN, Buckler EJ, et al. The Aim2Be mHealth Intervention for Children With Overweight or Obesity and Their Parents: Person-Centered Analyses to Uncover Digital Phenotypes. J Med Internet Res. 2022;24(6):e35285.
[8] Skjåkødegård HF, Conlon RPK, Hystad SW, et al. Family-based treatment of children with severe obesity in a public healthcare setting: results from a randomized controlled trial. Clin Obes. 2022;12(3)

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230318

Reliability of IMU-Derived Gait Parameters in Foot Drop Patients

Armando COCCIA[a,b,1], Federica AMITRANO[a,b], Gaetano PAGANO[a], Arcangelo BIANCARDI[a], Giuseppe TOMBOLINI[c] and Giovanni D'ADDIO[a]
[a] *Scientific Clinical Institutes ICS Maugeri SPA SB, Pavia, Italy*
[b] *University of Naples 'Federico II', Naples, Italy*
[c] *Officine Ortopediche Tombolini, Taranto, Italy*

Abstract. Foot drop is a deficit in foot dorsiflexion causing difficulties in walking. Passive ankle-foot orthoses are external devices used to support the drop foot improving gait functions. Foot drop deficits and therapeutic effects of AFO can be highlighted using gait analysis. This study reports values of the major spatiotemporal gait parameters assessed using wearable inertial sensors on a group of 25 subjects suffering from unilateral foot drop. Collected data were used to assess the test-retest reliability by means of Intraclass Correlation Coefficient and Minimum Detectable Change. Excellent test-retest reliability was found for all the parameters in all walking conditions. The analysis of Minimum Detectable Change identified the gait phases duration and the cadence as the most appropriate parameters to detect changes or improvements in subject gait after rehabilitation or specific treatment.

Keywords. Foot Drop, Ankle-Foot Orthosis, Gait Analysis, Inertial Measurement Unit, Intraclass Correlation Coefficient.

1. Introduction

Foot drop is a common deficit characterised by the difficulty in performing foot dorsiflexion, causing the front part of the foot to drag along the ground while walking. This severely affects gait functions: at heel strike, the forefoot generally impacts to the ground in an uncontrolled and rapid manner; during foot swing, the inability to lift the front part of the foot causes the toes to drag on the ground with consequent high risk of stumbling and falling [1]. The Codivilla spring is an ankle foot orthosis (AFO) designed to enhance control in walking and postural tasks, improving the quality of life of people suffering from this pathology [2]. This orthosis is made of thermoplastic material with an L-shape, with a rigid sole supporting the foot and a posterior leaf attached to the calf. The structure acts as a spring, returning elastically the flexion forces imparted during the terminal stance phase loading of the ankle. The effects of using these orthoses are generally qualitatively detectable from the observation of subjects walking. However, using gait analysis methodologies, the biomechanics of patients' gait can be studied to understand whether and how the use of the orthosis improves their walking. Based on these studies, it is also possible to customise the orthosis for specific purposes [3].

[1] Corresponding Author: Armando Coccia, E-mail: armando.coccia@icsmaugeri.it.

In foot drop patients abnormal plantarflexion is detectable during gait phases [4], while the use of an AFO determines improvements in walking speed, step frequency, stride and step length and functional ambulation ability [5–7]. In [8] significant differences were outlined between the affected and healthy limb in foot clearance at mid-swing and duration of the stance and swing phases. The step duration also significantly differed among limbs and improves when walking with AFO. While the effects of the orthotic devices on walking have been widely discussed in previous literature, it seems to lack the analysis of reliability of gait analysis metrics for this kind of deficit. Therefore, the aim of this study is to explore the intra-session test-retest reliability of the major spatio-temporal gait parameters measured on a cohort of foot drop patients using a system based on wearable Inertial Measurement Units (IMUs). Test-retest reliability represents a basic methodological study for clinical settings, as it validates the reliability of the measurement setting [9]. Moreover, the analysis of derived metrics, such as the Minimum Detectable Change (MDC), can also suggest which parameter or pattern is more suitable to characterize a specific pathological scenario.

2. Methods

2.1. Experimental Procedure

Twenty-five patients (17 males, 8 females), with unilateral foot drop syndrome (14 right, 11 left foot), were involved in this research (age 56.6 ± 13.5, BMI 23.3 ± 4.2). The Mobility Lab system by APDM (APDM Inc, Portland, OR, USA, http://apdm.com) was used to perform gait analysis on the study population. The experimental sessions were performed in the Movement Analysis Laboratory of the ICS Maugeri in Bari (Italy), and consisted of three repeated walking trials. Three body worn IMUs were used: one on the low back (just below L5 level) and two on the dorsal surface of the feet. The IMUs, measuring 43.7 x 39.7 x 13.7 mm (LxWxH), wireless transmit data sampled at 128Hz to an access point connected to the central workstation. In each trial the subject was instructed to stand quietly for 30s and then walk at comfortable speed over a 7m walkway, turn around a pivot and walk back to the starting point [10]. Subjects performed two separate sessions, in two different conditions: wearing or not the Codivilla spring on the affected limb. Both sessions were performed with patients wearing shoes. The order of the sessions was randomly selected, to avoid any ordering effects.

The following spatio-temporal parameters were considered in the analysis: foot clearance at mid-swing (cm), gait cycle time (GCT) (s), stance phase (expressed as percentage of GCT), swing phase (%GCT), cadence (steps/min), step duration (s), stride length (m), gait speed (m/s). These metrics were computed considering the gait cycles of each leg, thus producing two datasets for the affected and healthy (contralateral) limbs.

2.2. Statistical Analysis

Paired t-tests were conducted on the subject-averaged values to analyse whether statistically significant changes occur between limbs (affected vs contralateral) and/or walking conditions (walking with AFO or without).

The test-retest reliability over the three repeated trials was assessed using the Intraclass Correlation Coefficient (ICC) [11]. ICC estimates and their 95% confidence intervals (CIs) were calculated, on the four datasets per each spatio-temporal parameter,

based on single-measurement, absolute-agreement, 2-way mixed-effects model [12]. The ICC values were interpreted as poor when less than 0.5, moderate between 0.50 and 0.75, good between 0.75 and 0.90 and excellent when above 0.90 [13]. Absolute reliability was obtained using the Minimum Detectable Change (MDC), i.e. the smallest difference needed between separate measures on a subject to be considered a real change. It is calculated from the Standard Error of Measurement (SEM) and ICC as follow [14]: $SEM = SD_{all\ testing\ scores} \cdot \sqrt{(1 - ICC)}$ and $MDC = SEM \cdot 1.96 \cdot \sqrt{2}$. Statistical analyses were performed using R version 4.0.3 (R Foundation, Vienna, Austria).

3. Results

Table 1 reports descriptive statistics of the datasets in terms of mean and standard deviation, and results of paired t-tests. Statistically significant differences between limbs are indicated with an asterisk, while variations between walking conditions with a dagger symbol ($p - value < 0.05$). The test-retest reliability analysis produced the ICC values reported in Table 1. Excellent reliability ($ICC > 0.90$) was found for all the analysed parameters, except for the foot clearance in the healthy limb in walking without the AFO ($ICC = 0.865$). Table 2 also reports MDC values in the unit of the parameter and in percentage of the mean value to facilitate interpretation and comparisons. Low MDC values were found for stance and swing phases and cadence ($MDC\% < 10\%$). Gait cycle time, step duration, stride length and gait speed presented higher values of MDC, but lower than 15% of the mean value. Conversely, very high variability was underlined in foot clearance values over the trials.

4. Discussion and Conclusions

The principal aim of this study was to assess the test-retest reliability of gait analysis metrics on repeated measures performed with a system using three wearable inertial sensors. In the presented results, very high reliability was found for all the analysed parameters. The lowest ICC was found in the measures of healthy foot clearance at midswing when not wearing the AFO. These results showed that the gait analysis system based on three wearable inertial sensors, placed on the low back and on both feet, provides very reliable measures. This is in line with other scientific literature examining the performances of inertial sensors for gait analysis. In [15] high ICC values were found for gait metrics calculated using the APDM Mobility Lab with sensors placed on the ankle. Washabaugh *et al.* [14] found even higher ICC values when the system is used with the wearable sensors on the feet. The high repeatability of gait measures, obtained with inertial sensors, was confirmed also on other groups of patients [16,17] and with different systems and sensor configurations [18–20]. This study also explored the MDC for spatio-temporal gait parameters. The MDC was very low for stance, swing and cadence. Moderate values were found for step duration, stride length, gait cycle time and gait speed, while foot clearance MDC presented very high values. The values are consistent with those proposed by Washabaugh *et al.* [14], found on a group of 39 healthy subjects with analogous instrumentation, however authors did not explore foot clearance at midswing. Results about gait speed are similar to or better than those assessed with other methods [21].

Table 1. Descriptive statistics as mean ± standard deviation. Significant p-value from paired t-test between limbs are indicated with asterisk, while variations in walking conditions with.

	Without AFO		With AFO	
	Affected Foot	**Contralateral Foot**	**Affected Foot**	**Contralateral Foot**
Foot Clearance (cm)	2.29 ± 2.21	1.69 ± 0.77	2.40 ± 1.79	1.69 ± 0.78
Gait Cycle Time (s)	1.48 ± 0.39	1.49 ± 0.39	1.45 ± 0.37	1.45 ± 0.37
Stance (%GCT)	64.8 ± 4.3	67.2 ± 6.2*	64.5 ± 6.2	67.1 ± 5.6*
Swing (%GCT)	35.2 ± 4.3	32.8 ± 6.2*	35.6 ± 4.2	32.9 ± 5.6*
Cadence (steps/min)	85.7 ± 18.2	85.6 ± 18.2	87.0 ± 17.1	87.0 ± 17.2
Step Duration (s)	0.77 ± 0.21	0.72 ± 0.19*	0.73 ± 0.18†	0.72 ± 0.19
Stride Length (m)	0.84 ± 0.25	0.85 ± 0.23	0.85 ± 0.22	0.86 ± 0.22
Gait Speed (m/s)	0.63 ± 0.27	0.64 ± 0.27	0.65 ± 0.24	0.65 ± 0.25

Note: step duration, stride length and gait speed means have two significant digits because of gait analysis software limitation

Table 2. Results of test-retest reliability analysis. ICC values are reported with the 95% CI in parentheses. MDC values are expressed in the same unit as the gait parameter and in percentage of the mean value.

	ICC (95%CI)			
	Without AFO		**With AFO**	
	Affected Foot	**Contralateral Foot**	**Affected Foot**	**Contralateral Foot**
Foot Clearance	0.972(0.945; 0.987)	0.865(0.760; 0.933)	0.979(0.960; 0.990)	0.932(0.870; 0.967)
GCT	0.982(0.962; 0.992)	0.979(0.957; 0.990)	0.979(0.957; 0.990)	0.975(0.950; 0.989)
Stance	0.965(0.933; 0.983)	0.990(0.980; 0.995)	0.966(0.935; 0.984)	0.983(0.967; 0.992)
Swing	0.965(0.933; 0.983)	0.990(0.980; 0.995)	0.966(0.935; 0.984)	0.983(0.967; 0.992)
Cadence	0.981(0.959; 0.991)	0.980(0.958; 0.991)	0.982(0.963; 0.992)	0.982(0.963; 0.991)
Step Duration	0.979(0.958; 0.990)	0.969(0.941; 0.985)	0.976(0.952; 0.988)	0.965(0.928; 0.984)
Stride Length	0.984(0.969; 0.993)	0.985(0.959; 0.994)	0.974(0.949; 0.988)	0.977(0.951; 0.989)
Gait Speed	0.985(0.962; 0.993)	0.984(0.952; 0.994)	0.979(0.956; 0.991)	0.982(0.960; 0.992)
	MDC (MDC%)			
Foot Clearance	1.02(43.7%)	0.786(43.7%)	0.718(29.9%)	0.564(31.1%)
GCT	0.145(9.82%)	0.157(10.5%)	0.147(10.1%)	0.169(11.1%)
Stance	2.23(3.45%)	1.72(2.57%)	2.16(3.35%)	2.00(2.98%)
Swing	2.23(6.33%)	1.72(5.26%)	2.16(6.07%)	2.00(6.10%)
Cadence	6.99(8.15%)	7.15(8.35%)	6.40(7.36%)	6.48(7.44%)
Step Duration	0.0859(11.2%)	0.0909(12.6%)	0.0801(10.9%)	0.0985(13.7%)
Stride Length	0.0850(10.1%)	0.0799(9.38%)	0.0955(11.2%)	0.0934(10.9%)
Gait Speed	0.0924(14.6%)	0.0932(14.6%)	0.0968(15.0%)	0.0931(14.4%)

MDC is the smallest difference needed between separate measures to be considered a real change, so it represents a sort of threshold to assess the effects of rehabilitation. Changes in gait parameters lower than the MDC could be interpreted as measurement errors. This suggests that all the analysed parameters are suitable metrics to evaluate the pathological condition of the subject and to quantify the changes and improvements over time. Conversely, clinicians and researchers should cautiously interpret the variations in foot clearance, considering the variability of the parameter measured by the system, to verify whether it can be considered a relevant improvement for the subject. This confirms the poor discriminative value of the foot clearance metric, as discussed in [22] with a machine learning approach. Some limitations of the work should be mentioned: the study population is limited to 25 patients, a bigger cohort can provide more reliable results; the aetiology of foot drop is varied, this does not allow general assumptions about the cause of the deficit; moreover, patients had been using the orthosis for different periods of time, with different levels of confidence, this can have produced biases in the results. In future works the cohort of analysed subjects will be extended leading to more reliable analyses, which may be also detailed for each single pathology causing the walking deficit.

References

[1] Stewart JD. Foot drop: where, why and what to do? Practical Neurology. 2008 Jun;8(3):158–69.
[2] Bardelli R, Harlaar J, Morone G, Tomba P, Esquenazi A, Benedetti MG. The Codivilla spring: from then to now and beyond. European Journal of Physical and Rehabilitation Medicine. 2021 Dec;57(6).
[3] Coccia A, Amitrano F, Pagano G, Dileo L, Marsico V, Tombolini G, D'Addio G. Biomechanical modelling for quantitative assessment of gait kinematics in drop foot patients with ankle foot orthosis. In International Symposium on Medical Measurements and Applications (MeMeA). IEEE; 2022.
[4] Wiszomirska I, Błażkiewicz M, Kaczmarczyk K, Brzuszkiewicz-Kuźmicka G, Wit A. Effect of Drop Foot on Spatiotemporal, Kinematic, and Kinetic Parameters during Gait. Applied Bionics and Biomechanics. 2017 Apr;2017:1–6.
[5] Abe H, Michimata A, Sugawara K, Sugaya N, Izumi SI. Improving Gait Stability in Stroke Hemiplegic Patients with a Plastic Ankle-Foot Orthosis. The Tohoku Journal of Experimental Medicine. 2009;218(3):193–9.
[6] Farmani F, Mohseni-Bandpei MA, Bahramizadeh M, Aminian G, Abdoli A, Sadeghi-Goghari M. The Influence of Rocker Bar Ankle Foot Orthosis on Gait in Patients with Chronic Hemiplegia. Journal of Stroke and Cerebrovascular Diseases. 2016 Aug;25(8):2078–82.
[7] Choo YJ, Chang MC. Effectiveness of an ankle–foot orthosis on walking in patients with stroke: a systematic review and meta-analysis. Scientific Reports. 2021 Aug;11(1):15879–15879.
[8] Amitrano F, Coccia A, Pagano G, Dileo L, Losavio E, Tombolini G, D'Addio G. The Impact of Ankle-Foot Orthoses on Spatio-Temporal Gait Parameters in Drop-Foot Patients. In International Symposium on Medical Measurements and Applications (MeMeA). IEEE; 2022.
[9] Amitrano F, Coccia A, Donisi L, Pagano G, Cesarelli G, D'Addio G. Gait Analysis using Wearable E-Textile Sock: an Experimental Study of Test-Retest Reliability. In International Symposium on Medical Measurements and Applications (MeMeA). IEEE; 2021.
[10] Laurie King MM. Mobility Lab to Assess Balance and Gait with Synchronized Body-worn Sensors. Journal of Bioengineering and Biomedical Sciences. 2013;
[11] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods. 1996 Mar;1(1):30–46.
[12] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. 2016 Jun;15(2):155–63.
[13] Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. International Journal of Nursing Studies. 2011 Jun;48(6):661–71.
[14] Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C. Validity and repeatability of inertial measurement units for measuring gait parameters. Gait & Posture. 2017 Jun;55:87–93.
[15] Donisi L, Pagano G, Cesarelli G, Coccia A, Amitrano F, D'Addio G. Benchmarking between two wearable inertial systems for gait analysis based on a different sensor placement using several statistical approaches. Measurement. 2021 Mar;173:108642–108642.
[16] Coccia A, Lanzillo B, Donisi L, Amitrano F, Cesarelli G, D'Addio G. Repeatability of Spatio-Temporal Gait Measurements in Parkinson's Disease. In International Symposium on Medical Measurements and Applications (MeMeA) IEEE; 2020
[17] Coccia A, Amitrano F, Balbi P, Donisi L, Biancardi A, D'Addio G. Analysis of Test-Retest Repeatability of Gait Analysis Parameters in Hereditary Spastic Paraplegia. In International Symposium on Medical Measurements and Applications (MeMeA) IEEE; 2021.
[18] Henriksen M, Lund H, Moe-Nilssen R, Bliddal H, Danneskiod-Samsøe B. Test–retest reliability of trunk accelerometric gait analysis. Gait & Posture. 2004 Jun;19(3):288–97.
[19] De Ridder R, Lebleu J, Willems T, De Blaiser C, Detrembleur C, Roosen P. Concurrent Validity of a Commercial Wireless Trunk Triaxial Accelerometer System for Gait Analysis. Journal of Sport Rehabilitation. 2019 Aug;28(6).
[20] Fernández-Gorgojo M, Salas-Gómez D, Sánchez-Juan P, Barbado D, Laguna-Bercero E, Pérez-Núñez MI. Clinical–Functional Evaluation and Test–Retest Reliability of the G-WALK Sensor in Subjects with Bimalleolar Ankle Fractures 6 Months after Surgery. Sensors. 2022 Apr;22(8):3050–3050.
[21] Lewek MD, Randall EP. Reliability of Spatiotemporal Asymmetry During Overground Walking for Individuals Following Chronic Stroke. Journal of Neurologic Physical Therapy. 2011 Sep;35(3):116–21.
[22] Amitrano F, Coccia A, Cesarelli G, Donisi L, Pagano G, Cesarelli M, D'Addio G. The impact of ankle-foot orthosis on walking features of drop foot patients. In International Symposium on Medical Measurements and Applications (MeMeA)IEEE; 2022. p. 87–92.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230319

On the Difficulty of Predicting Engagement with Digital Health for Substance Use

Franziska GÜNTHER[a,1], Christopher YAU[b,c] , Sarah ELISON-DAVIES[d] and David WONG[a]
[a] *Centre for Health Informatics, School of Health Sciences, University of Manchester, UK*
[b] *Nuffield Department of Women's and Reproductive Health, University of Oxford, UK*
[c] *Health Data Research, UK*
[d] *LifeWorks, Canada*
ORCiD ID: Franziska Günther https://orcid.org/0000-0002-6016-765X

Abstract. Digital interventions can be an important instrument in treating substance use disorder. However, most digital mental health interventions suffer from early, frequent user dropout. Early prediction of engagement would allow identification of individuals whose engagement with digital interventions may be too limited to support behaviour change, and subsequently offering them support. To investigate this, we used machine learning models to predict different metrics of real-world engagement with a digital cognitive behavioural therapy intervention widely available in UK addiction services. Our predictor set consisted of baseline data from routinely-collected standardised psychometric measures. Areas under the ROC curve, and correlations between predicted and observed values indicated that baseline data do not contain sufficient information about individual patterns of engagement.

Keywords. prediction, digital health, substance use, engagement

1. Introduction

Digital interventions (DIs) for people with substance use disorders (SUDs) are digitalised complements or temporary replacements of traditional face-to-face therapies such as cognitive behavioural therapy (CBT). With DIs being more cost-effective and 24/7 accessible, they can represent an important instrument in treating SUDs.

To derive improved mental health outcomes via a DI, users need to engage with DI content to a sufficient degree. However, maintaining user engagement has been a consistent problem for DIs for mental health [1]. Early, accurate prediction of level of DI engagement could allow users at high risk of poor engagement to be identified. This could potentially be used to target additional support. The basis of such prediction should be data collected early into the user journey, if feasible at first user contact with a DI, as dropout after first use is a common phenomenon. However, it is not clear if prediction is at all possible using such data, since real-world engagement may depend on multiple factors that may not be reflected in a one-off clinical assessment before engagement.

The aim of this study is hence to assess whether engagement with one DI called Breaking Free Online (BFO) can be predicted using data collected at first DI use.

[1] Corresponding Author: Franziska Günther, E-Mail: franziska.gunther@postgrad.manchester.ac.uk.

2. Methods

2.1. Source of data

Data were routinely collected from BFO users enrolled between July 2016 and October 2022 in 513 community addiction services (CAS) in the UK. BFO is a self-guided digital CBT programme for SUDs, which for the past decade has been widely available to clients of CAS in the UK. Ethical approval for collection, storage and use of data accumulating from routine use of BFO by clients in participating treatment services, was obtained from an NHS Research Ethics Committee (London - South East, 16 May 2012 and 22 May 2017, references 12/LO/0076 and 12/LO/0287).

The BFO programme features six modules, with each split into one part psychoeducation and one part practice (applying what was learned in psychoeducation to one's own life). These subparts are subsequently referred to as "strategies", specifically, *information strategies* and *action strategies*.

Users are required to complete a baseline assessment so that modules can later be recommended to them. The baseline assessment includes four questionnaires: the Severity of Dependence Scale (SDS) [2], the Patient Health Questionnaire 4 [3], the World Health Organization Quality of Life measure (items 1, 2, 17, 18, and 20) [4] and the Recovery Progression Measure [5]. In addition, the baseline assessment also recorded user age, gender, ethnicity, abused substances, substance-using days in the preceding week and the user's target for substance-free days per week.

In addition to the assessment questionnaire data, assessment dates as well as module completion data, specifically the number of *information* and *action strategy* completions for each programme module and their most recent completion dates, were also available.

2.2. Predictors

The feature set we used to develop a prediction model of BFO engagement comprises 62 features corresponding to every question in the baseline assessment and a set of derived variables, specifically (1) baseline abstinence defined as zero substance-using days per week, (2) the number of days from registration to first assessment completion, (3) the number of clinical complexity inducing factors present at baseline (including financial difficulties, cravings, difficulties with physical health, at work, or with housing) and (4) cutoff-based variables on baseline anxiety, depression and substance dependence.

2.3. Outcomes

We used 9 continuous variables as outcomes with each measuring a different aspect of user engagement, as follows: (1) the number of days from the first to the last use event, subsequently referred to as the number of accessed days, (2) the number of strategies completed, (3) the number of *information strategies* completed, (4) the number of *action strategies* completed, (5) the number of use events (all assessments + strategies completed), (6) the use rate (number of use events / number of accessed days), (7) the percentage of days actively engaged (with the number of days on which an assessment was completed - which empirically fall together with known days of module completion in 67% percent of cases - regarded as active engagement), (8) the median intermission length in days (with days on which no active engagement was registered described as intermission days) and (9) the mean absolute deviation (MAD) intermission length. Log-

transformation was applied to all these variables. We also used the completion of 8 or more strategies as a binary outcome, referring to 8 sessions as the dose of talking therapy commonly received by patients completing a course of treatment through the NHS [6].

2.4. Statistical analysis and missing data

We predicted the 9 continuous outcomes and 1 binary outcome independently, using random forests and the XGBoost algorithm out-of-the-box with 10-fold cross validation. Stratification was applied to the target variable, with numeric strata being binned into quartiles. The average area under the receiver operating curve was used as a measure of predictive performance for binary outcome variables. Correlations between the observed and predicted values served as an assessment of predictive performance for the continuous outcome variables. The average root mean squared error (RMSE) was used to compare predictive performance between random forests and the XGBoost algorithm. We removed data from users who had >80% data missing on their baseline assessment (n = 706) as multiple imputation would be difficult for these users. For the remainder of the data, we opted for a complete case analysis as only 5% of these cases had incomplete data, and < 4% of cells were missing in total. R code for this analysis is available at https://github.com/franziskagunther/predict-engagement.

Table 1. User characteristics at baseline.

Characteristic	Statistic/Label	Value
Age in years	mean (SD, range)	40.1 (11.7, 18 - 84)
Gender	Female	47.1% (10,745)
	Male	52.5% (11,967)
	Other	0.3% (79)
Ethnicity	White	93% (21,207)
	Asian / Asian British	1.9% (426)
	Black / Black British	1.7% (382)
	Mixed	2.7% (626)
	Other	0.7% (150)
Primary substances	Alcohol	63.8% (14,533)
	Cocaine	11.7% (2,659)
	Marijuana	7.9% (1,810)
	Heroin	5.5% (1,248)
	Crack	3.5% (805)
	Other (46 other substances)	7.6% (1,741)
Substance dependence (SDS sum score equal to or larger than 3)	Yes	92.6% (20,494)
	No	7.4% (1,631)
Anxiety (sum of first two PHQ-4 items equal to or larger than 3)	Yes	69% (15,593)
	No	31% (7,007)
Depression (sum of last two PHQ-4 items equal to or larger than 3)	Yes	66.5% (15,023)
	No	33.5% (7,577)
Substance-using days in the past week	modes	0: 24.5%, 7: 38.1%

3. Results

We removed users who were younger than 18 (n = 82) or older than 89 years (n = 3). We also excluded users reporting a goal of increasing their substance consumption (n = 1,314, possibly due to user interpretation of item as the desired number of substance-consuming instead of substance-free days). Finally, we excluded users whose reports of

daily substance consumption was deemed to be clinically infeasible (n = 92). The final dataset included 22,796 users. Table 1 summarises their baseline characteristics.

We first examined individual feature-outcome correlations and found these to be low (see Figure 1). Cross-feature correlations were high, instead. Finally, we used XGBoost and random forests to see if the combination of predictors could predict outcomes, but predictive performance was poor in all cases. We report prediction accuracy with the random forests which performed slightly better than XGBoost with regards to RMSE and AUC. We obtained an average AUC of 0.57 [CI: 0.56-0.58] for the prediction of completing $n>=8$ modules. Model performance did not improve for other values of n. Predictive performance for continuous outcomes was similarly low and correlations between observed and predicted outcomes ranged between 0.03 and 0.13.

4. Discussion

Prediction of real-world engagement in self-guided DIs for mental health could contribute to overcoming one of the field's biggest problems; early and frequent dropout. Many DIs routinely administer assessments on users' clinical characteristics before providing access to DI content which, in theory, represent easily obtained sets of predictors of possibly non-beneficial engagement at the earliest possible time point.

We conducted a prediction study with data from the BFO programme in which all users, regardless of their actual level of engagement with the system, were included in the analysis. State-of-the-art prediction models were unable to accurately predict a range of engagement metrics from baseline assessment data which represents evidence against the predictability of engagement with BFO from clinical information at first access.

Multiple unmeasured factors may make prediction of user engagement challenging, such as the clinical complexity of individuals with SUD which likely interferes with engagement. Given the lack of prediction accuracy in this study, triaging new CAS clients for BFO use on the basis of their baseline assessment data may exclude individuals who may engage and possibly benefit from BFO if introduced to it.

This research has some limitations. First, our metrics of engagement were behavioural, and do neither reflect cognitive or emotional involvement of users with the BFO programme nor benefit which may be achieved after minimal engagement. However, by including continuous engagement variables, we attempted to reflect that beneficial engagement can have individually different outlooks which is often ignored in studies using "minimal engagement" thresholds. Our decision of allowing for a variety of different engagement patterns also resulted in intentionally not removing engagement outliers, which may bias our outcomes.

This study focused on a single DI. Examination of other DIs is desirable but complicated by access to commercially-sensitive data for independent researchers.

5. Conclusion

Early prediction of engagement is desirable in digital mental health. Our case study of prediction modeling of engagement in digital CBT for SUD suggests that information beyond clinical baseline characteristics is necessary to achieve accurate predictions.

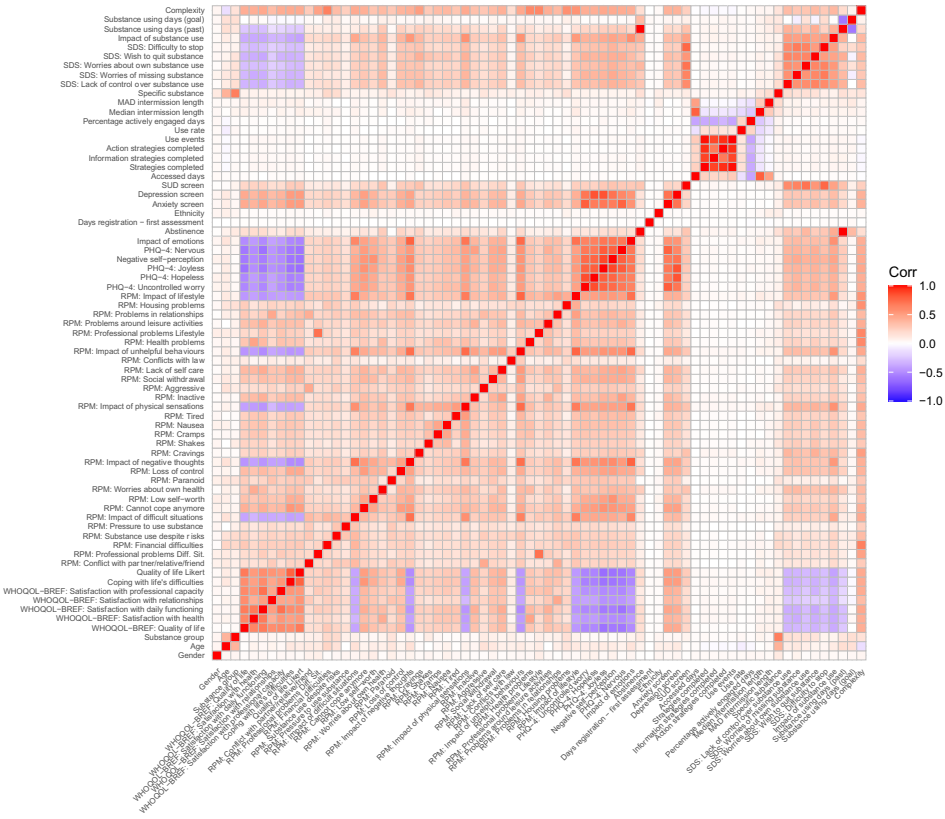


Figure 1. Correlations between variables used for prediction modelling.

References

[1] Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health. 2018 Aug;21(3):116-9, doi: 10.1136/eb-2018-102891

[2] Gossop M, Darke S, Griffiths P et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction. 1995 May;90(5):607-14, doi: 10.1046/j.1360-0443.1995.9056072.x

[3] Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ–4. Psychosomatics. 2009 Nov-Dec;50(6):613-21, doi: 10.1176/appi.psy.50.6.613

[4] Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL Group. Quality of Life Research. 2004 Mar;13(2):299-310, doi: 10.1023/B:QURE.0000018486.91360.00

[5] Elison S, Davies G, Ward J. Initial development and psychometric properties of a new measure of substance use disorder "recovery progression": the Recovery Progression Measure (RPM). Substance Use & Misuse. 2016 Jul;51(9):1195-206, doi: 10.3109/10826084.2016.1161052

[6] Population Health, Clinical Audit and Specialist Care Team, NHS Digital. Psychological therapies, annual report on the use of IAPT services, 2021-22 [Internet]. Leeds: NHS Digital; 2022 Sep 29. Report No. 12: Available from: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2021-22

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230320

The Assessment of Glioblastoma Metabolic Activity via ^{11}C-Methionine PET and Radiomics

Gleb DANILOV[a,1], Diana KALAEVA[b], Nina VIKHROVA[b], Tatiana KONAKOVA[b], Anna ZAGORODNOVA[a], Angelina POPOVA[a], Andrey POSTNOV[b], Svetlana SHUGAY[b], Michael SHIFRIN[a] and Igor PRONIN[b]
[a] *Laboratory of Biomedical Informatics and Artificial Intelligence, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*
[b] *Neuroimaging Department, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*
[c] *National Research Nuclear University MEPhI, Moscow, Russian Federation*
[d] *Lebedev Physical Institute of the Russiam Academy of Science, Moscow, Russian Federation*
[e] *JSC Research Institute of Technical Physics and Automation, Moscow, Russian Federation*
ORCiD ID: Gleb Danilov https://orcid.org/0000-0003-1442-5993

Abstract. Nowadays, the quantitative analysis of PET/CT data in patients with glioblastoma is not strictly standardized in the clinic and does not exclude the human factor. This study aimed to evaluate the relationship between the radiomic features of glioblastoma ^{11}C-methionine PET images and the tumor-to-normal brain (T/N) ratio determined by radiologists in clinical routine. PET/CT data were obtained for 40 patients (mean age 55 ± 12 years; 77.5% men) with a histologically confirmed diagnosis of glioblastoma. Radiomic features were calculated for the whole brain and tumor-containing regions of interest using the RIA package for R. We redesigned the original RIA functions for GLCM and GLRLM calculation to reduce computation time significantly. Machine learning over radiomic features was applied to predict T/N with the best median correlation between the true and predicted values of 0.73 ($p = 0.01$). The present study showed a reproducible linear relationship between ^{11}C-methionine PET radiomic features and a T/N indicator routinely assessed in brain tumors. Radiomics enabled utilizing texture properties of PET/CT neuroimaging that may reflect the biological activity of glioblastoma and can potentially augment the radiological assessment.

Keywords. Glioblastoma, radiomics, neuroradiomics, PET, machine learning, artificial intelligence, neurosurgery

[1] Corresponding author: Gleb Danilov, N.N. Burdenko Neurosurgery Center, 4th Tverskaya-Yamskaya str. 16, Moscow 125047, Russian Federation; E-mail: glebda@yandex.ru.

1. Introduction

Glioblastoma is the most common primary malignant astrocytic neoplasm of the brain [1]. A valuable neuroimaging modality in glioblastoma diagnostics is PET/CT with radiopharmaceutical. The intensity of its uptake in a tumor relative to a normal brain (tumor-to-normal brain ratio, T/N) correlates with the biological aggression of the neoplasm. To date, the quantitative analysis of PET/CT data in patients with glioblastoma is not strictly standardized in the clinic and does not exclude the human factor. Radiomics can promote unification and increase the objectivity and informativeness of neuroimaging assessment [2]. This study aimed to evaluate the relationship between the radiomic features of glioblastoma ^{11}C-methionine PET images and the T/N ratio determined by radiologists in clinical routine. We believe the proposed approach facilitates the standardization and automation of radiological biomarkers production in clinical practice.

2. Methods

Our observational study was conducted under the ethical principles in the Helsinki Declaration of the World Medical Association (1964). We obtained PET images from adult patients with supratentorial glioblastoma treated at the National Medical Research Center for Neurosurgery named after N.N. Burdenko between 2018 and 2020. To assess the relative metabolic activity of ^{11}C-methionine in glioblastoma by PET as a clinical routine, the average values of the standardized uptake value (SUV) were calculated in 1.0 cm^3 of the most active tumor region (SUVt) and 1.0 cm^3 of normal brain tissue of the contralateral frontal lobe (SUVn). Then the tumor-to-brain ratio was derived as T/N = SUVt/SUVn.

To calculate the radiomic features, MRI and PET/CT were co-registered using the PMOD software (v 4.0). Voxel values outside of the head contours were eliminated. A fixed-size rectangular area of interest was set for all the co-registered slices to capture the maximum tumor volume on any level. Thus, in a 3D space, the entire tumor was enclosed in a parallelepiped. Then, for each patient, the whole 3D array of PET voxels ("whole brain" dataset, WBD) and a subset of the 3D array of PET voxels in a given parallelepiped ("cropped brain" dataset, CBD) were exported to separate NIfTI files, which were used to compute radiomic features.

Calculations and data analysis were performed using the R programming language (version 4.2.2) in the RStudio Server IDE (version 2022.07.0+548) on an NVIDIA DGX A100 supercomputer. Radiomic features were computed from PET 3D array using the *RIA* library [3]. The voxel values from CBD were discretized into 2, 4, 8, 16, 32, 64, and 128 bins. WBD was discretized only into 128 levels to reduce the time and computation burden. We calculated first-order, gray level co-occurrence matrix (*GLCM*), gray level run length matrix (*GLRLM*) and geometry-based statistics (the complete list of features is presented in [4]). To compute PET radiomic features for the entire 3D brain image, we redesigned the original functions from the *RIA* package that calculate *GLCM* and *GLRLM*. That accelerated the computations a thousandfold and enabled the whole brain radiomics with a reasonable amount of time.

At the first step of data analysis, we selected radiomic features showing statistically significant Pearson correlation with the T/N ratio ($p < 0.05$). Then the linear regression models with LASSO regularization were trained over selected radiomics features as

predictors and T/N ratio as the target variable using *glmnet* library. The predictors were normalized, and the target variable was transformed using the decimal logarithm. Machine learning was repeated in 300 tests. The training and test samples were randomly split in each trial as 70% and 30% of the original dataset. The mean absolute error (MAE), root mean squared error (RMSE), and Pearson and Spearman correlation coefficients between the true and predicted T/N values were calculated in each test to evaluate prediction quality. The results from all tests were summarized to produce more robust estimates.

3. Results

A total of 40 independent preoperative PET/CT studies from 40 patients (31 (77.5%) men and 9 (22.5%) women, avg. age 55 ± 12 years) were included. The median T/N ratio obtained by neuroradiologists for PET studies was 3.26 [2.74; 4.17], and the minimum and maximum values were 1.94 and 5.03. MAE, RMSE, and the number of predictors (NoP) in models summarized from 300 tests as median [25% quantile; 75% quantile] and calculated exclusively on test samples, as well as the minimum (Min NoP) and the maximum (Max NoP) number of predictors, are presented in Table 1. The numbers in dataset names ("2_4..._128" or "_128") denote all discretization levels for images from which the radiomics parameters included in the dataset were calculated.

Table 1. The quality metrics of linear regression models summarized from 300 tests for various sets of radiomic features.

Dataset	MAE	RMSE	NoP	Min NoP	Max NoP
WBD_128_CBD_128	0.49 [0.41;0.58]	0.68 [0.54;0.85]	14 [11;16]	4	32
WBD_128_CBD_2_4_8_16_32_64_128	0.52 [0.45;0.63]	0.75 [0.62;0.91]	17 [14;20]	2	37
CBD_128	0.54 [0.46;0.64]	0.76 [0.63;0.94]	9 [7;12]	1	27
CBD_2_4_8_16_32_64_128	0.56 [0.47;0.67]	0.61 [0.45;0.74]	11 [9;15]	1	28
WBD_128	0.72 [0.63;0.81]	0.90 [0.79;1.02]	9 [6;12]	0	23

Table 2 shows the correlation between the predicted and true T/N ratio for linear regression models with regularization and its statistical significance calculated on test samples and presented as median [25% quantile; 75% quantile]. The rows in Tables 1 and 2 range from best (top) to worst (bottom) performance.

Table 2. Correlation between the predicted and true T/N ratios.

Dataset	Correlation	Coefficient	P-value
WBD_128_&_CBD_128	Pearson	0.71 [0.55;0.82]	0.01 [0.00;0.06]
WBD_128_&_CBD_128	Spearman	0.73 [0.55;0.81]	0.01 [0.00;0.07]
WBD_128_&_CBD_2_4_8_16_32_64_128	Pearson	0.67 [0.53;0.77]	0.02 [0.00;0.08]
WBD_128_&_CBD_2_4_8_16_32_64_128	Spearman	0.67 [0.51;0.80]	0.02 [0.00;0.09]
CBD_128	Pearson	0.64 [0.45;0.80]	0.03 [0.00;0.14]
CBD_128	Spearman	0.67 [0.48;0.80]	0.02 [0.00;0.12]
CBD_2_4_8_16_32_64_128	Pearson	0.61 [0.45;0.74]	0.03 [0.01;0.14]

CBD_2_4_8_16_32_64_128	Spearman	0.62 [0.47;0.78]	0.04 [0.00;0.12]
WBD_128	Pearson	0.36 [0.20;0.53]	0.24 [0.08;0.49]
WBD_128	Spearman	0.42 [0.28;0.56]	0.17 [0.06;0.38]

Figures 1A and 1C show the regression lines between true and predicted T/N ratios superimposed for 300 tests over WBD_128_CBD_128 dataset (best performance) and WBD_128 dataset (worst performance), respectively. Similarly, figures 1B and 1D present the locally weighted scatterplot smoothing (LOESS) lines for all the tests to catch the most common trends.

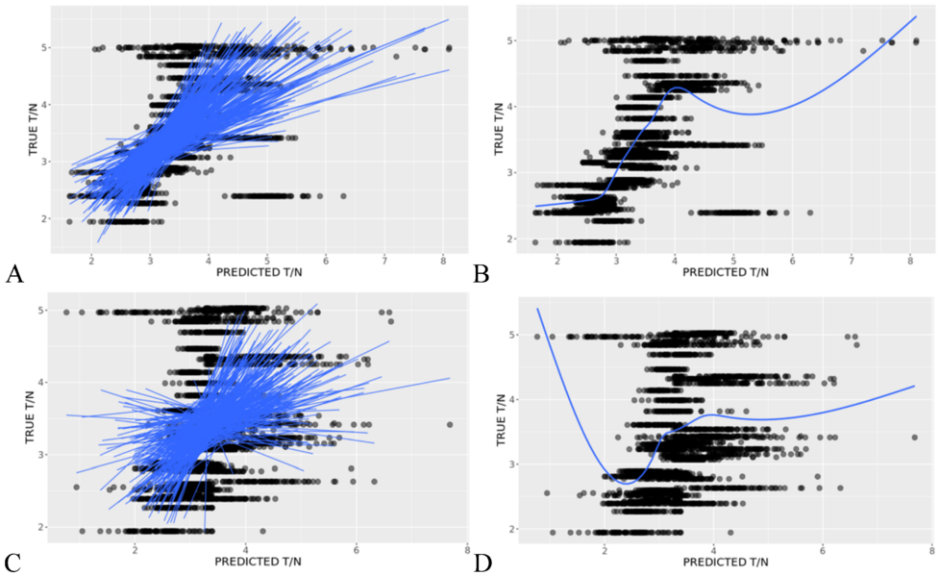


Figure 1. The scatterplot of predicted vs. true T/N ratios from 300 tests. A – superimposed regression lines, WBD_128_CBD_128 dataset; B – LOESS line, WBD_128_CBD_128 dataset; C – superimposed regression lines, WBD_128 dataset; B – LOESS line, WBD_128 dataset;

4. Discussion

The present study showed a reproducible linear relationship between ^{11}C-methionine PET radiomic features and a T/N indicator routinely assessed for brain tumors. In other words, we demonstrated the fundamental capability of calculating radiomics-based complex PET biomarkers that could be clinically relevant, e.g. as predictors of tumor proliferative activity. It is essential that these regularities were found for glioblastoma PET images - within one top-malignant histological tumor type. However, these effects should be tested within a comprehensive histological range.

The impact of whole brain radiomics is well-noted in Tables 1 and 2 and Figures 1. The WBD alone is less effective compared to sole CBD in predicting T/N. The combination of radiomic features from a set of cropped images with different discretization levels is slightly inferior to CBD with one top-discretization level. However, combining radiomic features from the entire brain and its tumor-containing area provided the best value. That strikes the importance of selecting the regions of

interest for radiomics (not limited to the visible tumor) and the possible effect of discretization, which should be further tested.

A common radiomics application in glioblastoma imaging studies is the differential diagnosis, overall survival prognosis, and molecular biomarkers prediction [5]. However, radiomics was rarely used to study PET/CT glioblastoma images [6]. The potential clinical significance of radiological PET biomarkers was previously shown in survival research for patients with glioblastoma [7]. The application of radiomics to ^{11}C-methionine PET is extremely rare. To the best of our knowledge, this is the first study to combine local and whole-brain radiomics to predict tumor metabolic activity by PET/CT with ^{11}C-methionine.

The main limitations of our study are the small sample size and the lack of standards for identifying the local region of interest in capturing the tumor. Our future work will address more radiomic features evaluation on larger samples.

Conclusion

Radiomics enables utilizing texture properties of PET/CT neuroimaging that may reflect the biological activity of glioblastoma and can potentially augment the radiological assessment. Despite the current limitations in the application, the first results indicate the promising potential of neuroradiomics. The regularities found in this research should be tested with a larger amount of data.

The study was supported by the Ministry of Science and Higher Education of the Russian Federation under agreement No. 075-15-2021-1343.

References

[1] Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. Neuro-oncology. 2021 Oct;23(Supplement_3):iii1-05.. doi:10.1093/NEUONC/NOAB200.

[2] Litvin AA, Burkin DA, Kropinov AA, Paramzin FN. Radiomics and digital image texture analysis in oncology. Современные технологии в медицине. 2021;13(2 (eng)):97-104. doi:10.17691/STM2021.13.2.11.

[3] RIA package manual, (n.d.). https://cran.r-project.org/web/packages/RIA/vignettes/RIA.html (accessed November 25, 2022).

[4] Kolossváry M, Karády J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B, Maurovich-Horvat P. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. Circulation: Cardiovascular Imaging. 2017 Dec;10(12):e006843..

[5] Zhu M, Li S, Kuang Y, Hill VB, Heimberger AB, Zhai L, Zhai S. Artificial intelligence in the radiomic analysis of glioblastomas: A review, taxonomy, and perspective. Frontiers in Oncology. 2022:3793. doi:10.3389/FONC.2022.924245.

[6] Martin P, Holloway L, Metcalfe P, Koh ES, Brighi C. Challenges in Glioblastoma Radiomics and the Path to Clinical Implementation. Cancers. 2022 Aug 12;14(16):3897. doi:10.3390/CANCERS14163897.

[7] Li Z, Holzgreve A, Unterrainer LM, Ruf VC, Quach S, Bartos LM, Suchorska B, Niyazi M, Wenter V, Herms J, Bartenstein P. Combination of pre-treatment dynamic [18F] FET PET radiomics and conventional clinical parameters for the survival stratification in patients with IDH-wildtype glioblastoma. European Journal of Nuclear Medicine and Molecular Imaging. 2023 Jan;50(2):535-45. doi:10.1007/S00259-022-05988-2/TABLES/3.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230321

Benchmarking the Impact of Noise on Deep Learning-Based Classification of Atrial Fibrillation in 12-Lead ECG

Theresa BENDER[a,b,1], Philip GEMKE[a], Ennio IDROBO-AVILA[a], Henning DATHE[a], Dagmar KREFTING[a,b] and Nicolai SPICHER[a,b]
[a] *Department of Medical Informatics, University Medical Center Göttingen, Göttingen, Germany*
[b] *DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Göttingen, Germany*

Abstract. Electrocardiography analysis is widely used in various clinical applications and Deep Learning models for classification tasks are currently in the focus of research. Due to their data-driven character, they bear the potential to handle signal noise efficiently, but its influence on the accuracy of these methods is still unclear. Therefore, we benchmark the influence of four types of noise on the accuracy of a Deep Learning-based method for atrial fibrillation detection in 12-lead electrocardiograms. We use a subset of a publicly available dataset (PTB-XL) and use the metadata provided by human experts regarding noise for assigning a signal quality to each electrocardiogram. Furthermore, we compute a quantitative signal-to-noise ratio for each electrocardiogram. We analyze the accuracy of the Deep Learning model with respect to both metrics and observe that the method can robustly identify atrial fibrillation, even in cases signals are labelled by human experts as being noisy on multiple leads. False positive and false negative rates are slightly worse for data being labelled as noisy. Interestingly, data annotated as showing baseline drift noise results in an accuracy very similar to data without. We conclude that the issue of processing noisy electrocardiography data can be addressed successfully by Deep Learning methods that might not need preprocessing as many conventional methods do.

Keywords. Deep Learning, Electrocardiogram, Atrial Fibrillation, Noise

1. Introduction

Electrocardiograms (ECGs) are recordings of the electrical activity of the heart and are frequently used in emergency and in-patient care. However, different types of noise, either stemming from the patient's behaviour (e.g. motion) or the devices (e.g. power line interference), can be introduced during measurement. The presence of noise leads to a twofold problem: It impedes detection of anomalies leading to false findings and alarms [1] and, if the signal-to-noise ratio (SNR) reaches a certain level, detecting diagnostically relevant features becomes impossible [2].

[1] Corresponding Author: Theresa Bender, University Medical Center Göttingen, Robert-Koch-Str. 40, 37075 Göttingen, Germany, E-mail: theresa.bender@med.uni-goettingen.de.

One class of features with high clinical importance are the so-called "fiducial points", i.e. the center, on- and offsets of ECG waves such as the QRS complex and the P-/T-wave. They are used for segmenting heartbeats into meaningful intervals [3] and by doing so allow for arrhythmia detection. Atrial fibrillation (AF) is the most prevalent arrhythmia which is characterized by uncoordinated electrical impulses in the atrium and might lead to severe cardiovascular issues, such as stroke or heart failure. Analyzing the interval in a heartbeat where a P-wave is expected is crucial for AF classification as its absence indicates a lack of sinoatrial node activity and is thereby a sign for AF [4]. However, so-called fibrillatory waves might occur, mimicking P-waves, impeding the assessment of sinoatrial node activity.

Many state-of-the-art algorithms for ECG classification are based on extracting semantic features derived from human expert knowledge, such as fiducial points. However, as these algorithms tend to wrong results in case of noise [5], various denoising strategies [6] have been proposed. In contrast, algorithms from the field of deep learning (DL) were explored for ECG classification tasks recently [7,8]. Instead of semantic features, they are based on agnostic features derived from fully-automatic correlation analysis between input ECGs and output classes in an end-to-end fashion. These models are based on the underlying premise that training and test datasets are stemming from the same distribution, which is often their pitfall in case of dataset shifts (variant devices, users, noise). Although initial studies indicate a better robustness to high SNRs [9,10], it remains unclear to which extend it affects these models.

Thereby, in this work we benchmark the accuracy of a state-of-the-art pre-trained DL model for 12-lead ECG classification regarding its susceptibility to different types of noise. We use the publicly available PTB-XL dataset which contains annotations for several categories of noise made by human technical experts and compare the model's accuracy w.r.t. type of noise.

2. Methods

We analyze a subset of the PTB-XL dataset containing 12-lead ECGs of 10 second length with a sampling rate of 500 Hz that were acquired between 1989 and 1996 [11]. The subset contains all 1,514 ECGs annotated as showing AF (label in PTB-XL: *AFIB*) and we add the first 2,000 normal ECGs (*NORM*) as healthy controls. For each signal, we use a qualitative and a quantitative method to estimate SNR.

2.1. SNR Based on Annotations (SNR_a)

For each ECG we determine the number of noisy leads using the columns *baseline_drift*, *static_noise*, *burst_noise* and *electrodes_problems* provided in the PTB-XL metadata. In the majority of cases, they contain the name of a single lead (e.g. "aVL"), multiple leads ("I,aVR") or ranges (e.g. "I-III"). Using a custom script, we convert this information to numeric values ranging from 0 to 12 for each type of noise. The labels "alles" (all) and "noisy recording" are converted to 12. We remove ECGs associated with other labels as they are of a more qualitative nature (e.g. "leicht" (light)). In this way, for each signal a qualitative, unit-less, linear SNR measure is computed, ranging from 0 (no noise reported) to 12*4=48 (all leads are affected by all types of noise). As shown in Tbl. 1, we use this information to split the dataset in ECGs without ("w/o") a noise label and ECGs with ("w/") a noise label.

Table 1. Properties of subset extracted from PTB-XL (left) and results of DL-based AF classification (right). ECGs are grouped according to annotations: In case there is one or more noise label in the metadata, an ECG is assigned to "w/", else to "w/o". FP and FN denote False Positive and False Negative, respectively.

Noise Label	AF	Healthy controls
w/o	1,097	1,581
w/	417	419

Noise Label	DL: FP	DL: FN
w/o	0.04 %	3.96 %
w/	0.24 %	7.06 %

It has to be underlined that a value of zero does not have to mean that there is no noise, it just reflects that there is a potential for a noise-free ECG. The authors of PTB-XL also indicated that missing annotations in case of artifacts or false annotations in case of noise-free signals might occur. However, they concluded that the metadata bears the potential for ECG quality assessment [12].

2.2. Measured SNR (SNR_m)

Due to the limitations of the manual annotations and as they are only available for 22 % of the PTB-XL database [12], we additionally use a quantitative SNR measure for each signal. We compute the Fourier Transform of the signals as well as the ratio of energies in two frequency bands as proposed in [13]. Based on the expected heart rates during AF, we define the "signal" frequency band ranging from 40 to 150 beats-per-minute (0.66 to 2.5 Hz) and define the "noise" frequency band as < 40 and > 150 beats-per-minute. By scaling with 10 log 10, we arrive at an SNR expressed in logarithmic decibel scale (dB).

2.3. DL Classification

ECG data is classified with a pre-trained model by Ribeiro et al. [7]. The model is a residual network and was trained on more than two million ECGs that were acquired within a Brazilian telehealth network. It outputs independent probabilities for six abnormalities, but we limit our analysis to AF. We use a threshold defined by the authors[2].

2.4. Data Analysis

We analyze the subset regarding differences between ECGs with and without noise labels for i) their distribution of SNR_m and SNR_a as well as ii) the accuracy of DL classification of each noise category. For ii) we compared the noisy recordings ($SNR_a > 0$) with randomly drawn signals from equally sized control groups ($SNR_a = 0$).

3. Results

Fig. 1 shows the distribution of SNR_a and SNR_m values on the left and right side. The majority of ECGs with noise labels has less than 15 with the maximum being 29. This shows that even in the duration of 10 seconds, different data quality issues per lead may occur. SNR_m values are occurring in the range of [-33.03,-7.78] dB with no clear difference between ECGs with and without noise labels.

[2] https://github.com/antonior92/automatic-ecg-diagnosis/blob/master/generate_figures_and_tables.py}, commit 89f929d, line 121

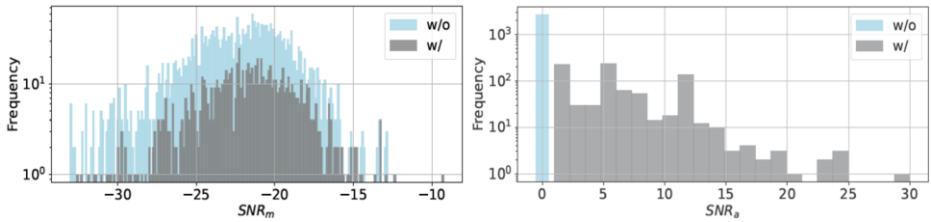


Figure 1. Distribution of values of both SNR metrics with (grey) and without (blue) noise labels.

Table 2. DL accuracy w.r.t. the four types of noise. The variable *n* represents the number of signals with the given label (w/). For comparison to signals without a label (w/o), *n* ECGs are randomly drawn 100 times and accuracy is given as mean ± standard deviation.

Label \ Type	**Baseline Drift** (*n* = 305)	**Static Noise** (*n* = 478)	**Burst Noise** (*n* = 156)	**Electrode Problems** (*n* = 6)
w/o	96.8 % ± 0.9 %	96.8 % ± 0.7 %	96.9 % ± 1.3 %	96.3% ± 8.0 %
w/	97.7 %	94.6 %	94.9 %	100.0 %

Tbl. 1 (right) shows FP and FN rates of AF classification w.r.t. the existence of noise labels. FP is worsened by 0.2 % and FN by 3.1 % in case ECGs are annotated with noise labels. Tbl. 2 shows the DL accuracy for each type of noise compared to the same number of ECGs but randomly drawn 100 times from data without noise labels. ECGs with baseline drift or electrode problems are classified more accurately in comparison to random ECG signals without noise annotations, whereas ECGs with annotated burst and static noise reveal worse performance.

4. Discussion

In general, the DL model robustly classifies AF, even in case ECGs are labelled by human experts as having multiple leads influenced by noise. Interestingly, in presence of baseline drift or electrode problems, accuracy is not deteriorated, but within one standard deviation compared to signals without noise labels. As a limitation, it has to be underlined that annotations are non-complete [12] and the subset contains only six signals annotated with electrode problems.

As the DL model can be assumed as a “black box”, we can only speculate about the reasons for this behaviour. It could be explained by partial misinterpretation of baseline drift or static noise as P-waves. As we could show in previous work [14], the DL model was trained such that P-waves and R-peaks have a high relevance, similar to human perception, while numerous other features influence its decision. This multi-factor decision process could be robust to different kinds of noise, but this requires its presence during training. A shift between training and test datasets is always an issue for DL models. To mitigate this effect is has been suggested to intentionally include noise during training [9]. The model used in this work was trained on two million non-public ECGs.

However, since the distribution of SNR_m looks visually similar with or without noise labels, SNR_a might not be optimal for quality assessment on its own. A “no noise” label, explicitly identifying ECGs without data quality issues, and more labels in general would be a valuable addition for future experiments.

5. Conclusion

Results show that the DL model is able to detect AF in 12-lead ECGs with high accuracy, even in the presence of data quality issues according to human experts. We conclude that end-to-end DL models based on agnostic features can address the difficulty of processing noisy ECGs. In contrast to conventional methods based on semantic features, they might not require preprocessing methods for achieving high accuracy. However, more experiments with larger and more diverse datasets should be the subject of future work.

References

[1] Festag S, Spreckelsen C. Semantic Anomaly Detection in Medical Time Series. Stud Health Technol Inform 2021; 278:118–25, doi: 10.3233/SHTI210059.
[2] Apandi ZFM, Ikeura R, Hayakawa S, Tsutsumi S. An Analysis of the Effects of Noisy Electrocardiogram Signal on Heartbeat Detection Performance. Bioengineering 2020; 7(2), doi: 10.3390/bioengineering7020053.
[3] Spicher N, Kukuk M. Delineation of Electrocardiograms Using Multiscale Parameter Estimation. IEEE J Biomed Health Inform 2020; 24(8):2216–29, doi: 10.1109/JBHI.2019.2963786.
[4] Kreimer F, Aweimer A, Pflaumbaum A, Mügge A, Gotzmann M. Impact of P-wave indices in prediction of atrial fibrillation-Insight from loop recorder analysis. Ann Noninv Electrocard 2021; 26(5):e12854, doi: 10.1111/anec.12854.
[5] Kumar P, Sharma VK. Detection and classification of ECG noises using decomposition on mixed codebook for quality analysis. Healthc Technol Lett 2020; 7(1):18–24, 10.1049/htl.2019.0096.
[6] Mir HY, Singh O. ECG denoising and feature extraction techniques – a review. Journal of Medical Engineering \& Technology 2021; 45(8):672–84, doi: 10.1080/03091902.2021.1955032.
[7] Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020; 11(1):1760, doi: 10.1038/s41467-020-15432-4.
[8] Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. Eur Heart J 2021; 42(46):4717–30, doi: 10.1093/eurheartj/ehab649.
[9] Venton J, Harris PM, Sundar A, Smith NAS, Aston PJ. Robustness of convolutional neural networks to physiological electrocardiogram noise. Philos Trans A Math Phys Eng Sci 2021; 379(2212):20200262, doi: 10.1098/rsta.2020.0262.
[10] Sraitih M, Jabrane Y, Hajjam El Hassani A. A Robustness Evaluation of Machine Learning Algorithms for ECG Myocardial Infarction Detection. JCM 2022; 11(17):4935, doi: 10.3390/jcm11174935.
[11] Wagner P, Strodthoff N, Bousseljot R-D, Kreiseler D, Lunze FI, Samek W et al. PTB-XL, a large publicly available electrocardiography dataset. Sci Data 2020; 7(1):154, doi: 10.1038/s41597-020-0495-6.
[12] Strodthoff N, Wagner P, Schaeffter T, Samek W. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL; 2020 Apr 28, doi: 10.48550/arXiv.2004.13701.
[13] Haan G de, Jeanne V. Robust pulse rate from chrominance-based rPPG. IEEE Trans Biomed Eng 2013; 60(10):2878–86, doi: 10.1109/TBME.2013.2266196.
[14] Bender T, Beinecke JM, Krefting D, Müller C, Dathe H, Seidler T et al. Analysis of a Deep Learning Model for 12-Lead ECG Classification Reveals Learned Features Similar to Diagnostic Criteria; 2022, doi: 10.48550/arXiv.2211.01738.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230322

Patient Perspectives on Long-Term Use of a Pulmonary Telerehabilitation Platform: A Qualitative Analysis

Aileen S. GABRIEL[a,1], Irena PARVANOVA[a] and Joseph FINKELSTEIN[a]
[a] *Center for Biomedical and Population Health Informatics, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
ORCiD ID: Aileen Gabriel https://orcid.org/ 0000-0002-5872-9078

Abstract. To effectively develop patient-centered interfaces and functionality, it is essential to investigate different viewpoints on pulmonary telerehabilitation. The purpose of this study is to explore the views and experiences of COPD patients after the completion of a 12-month home-based pulmonary telerehabilitation program. Semi-structured qualitative interviews were conducted with 15 COPD patients. The interviews were analyzed using a thematic analysis approach to deductively identify patterns and themes. Patients responded with approval for the telerehabilitation system, particularly for its convenience and ease of use. This study offers a thorough investigation of patient viewpoints when utilizing the telerehabilitation technology. These insightful observations will be considered for future development and implementation of a patient-centered COPD telerehabilitation system to provide support tailored to patient needs, preferences, and expectations.

Keywords. Telerehabilitation, chronic obstructive pulmonary disease, qualitative analysis

1. Introduction

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of morbidity and mortality worldwide. According to the World Health Organization, COPD will overtake heart disease as the third largest cause of death by 2023 [1]. There is strong data demonstrating that patients with COPD who participate in pulmonary rehabilitation experience improved functional ability, decreased dyspnea, and better clinical outcomes [2]. Traditional or center-based pulmonary rehabilitation (PR) programs involve weekly or biweekly in-person sessions at an outpatient clinic for 6–8 weeks though sustainable outcome improvement in COPD can be achieved only with life-long rehabilitation. Despite the advantages of center-based PR, adoption rates are still modest. Transportation difficulties, tight budgets, conflicting schedules, staff shortages, and lack of perceived benefits are barriers to participation [3]. Alternative PR delivery methods, such as home-based PR and telerehabilitation, have been shown to produce clinical effects that are equivalent to those of the typical in-person PR program offered at an outpatient clinic [4]. The results of previous studies evaluating the views of patients who underwent an 8-week home-based PR point to the program's high level of patient

[1] Corresponding Author: Aileen S. Gabriel, E-mail: Aileen.Gabriel@mssm.edu.

acceptance [4]. Patients also highly valued the flexibility and convenience of the home-based program. However, a thorough qualitative investigation of COPD patients' perceptions of long-term pulmonary telerehabilitation has not been conducted. The goal of this study is to explore the views and experiences of COPD patients who completed a 12-month home-based pulmonary telerehabilitation program.

2. Methods

The current study is part of a larger project with the goal of systematically assessing the impact of the Comprehensive Health Informatics Engagement Framework for Pulmonary Rehabilitation (CHIEF-PR) in a randomized controlled trial. The study was approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. Participants consisted of patients diagnosed with COPD who were hospitalized with COPD exacerbation at the Mount Sinai Health System in New York. Patients with an acute exacerbation of COPD within four weeks of enrollment were invited to participate in the study. Participants were recruited until thematic saturation when interviews generate few or no new information, ideas, themes, or codes [5]. Semi-structured qualitative interviews were conducted with 15 participants who completed a 12-month home-based pulmonary telerehabilitation program which was previously described [6]. The qualitative interviews consisted of open-ended questions examining the participant's experience with the content, interface, and the process of using a web-based telerehabilitation system called Home Automated Telemanagement (HAT) [6]. The HAT system was created to support healthcare practitioners in treating, educating, and monitoring patients with COPD and to help patients adhere to their personalized PR program according to current clinical guidelines. The interviews were conducted remotely by a trained researcher using videoconference software. Each interview session lasted approximately 20 minutes. The responses for each open-ended question were recorded in separate de-identified documents using Microsoft Word which serves as the raw data set.

The qualitative data were analyzed using a thematic analysis method which consists of a six-phase process to organize, identify, and interpret key patterns and themes [7]. As part of the first phase of thematic analysis known as data familiarization, researchers reviewed and then utilized Microsoft Excel to consolidate and organize the raw textual data across all participants using a framework analysis approach. This systematic approach is essentially a comparative type of thematic analysis which uses an a priori structure of concepts to organize the raw data [8]. The data was organized in alignment with the three usability areas evaluated in the semi-structured qualitative interviews: content, interface, and process. During the second phase, initial codes were generated using a deductive or 'top-down' approach where codes are strongly linked to the data and map onto a specific research question or theme [7]. The codes were aggregated to identify and summarize into general themes and sub-themes as part of phases three and four. In phase five, the themes were reviewed and refined with clear descriptions and examples to illustrate each theme. The sixth phase consists of reporting the findings of the thematic analysis and is included in the following results section.

3. Results

3.1. Themes Related to the Content of the Pulmonary Telerehabilitation System

Users were asked to share their thoughts about the exercise program content, including the instructions, self-report options, performance feedback, and demonstrations on how to perform the exercise. The responses indicated that the exercise content was well received. Users described the exercise instructions as sufficient, informative, and easy to follow. Users also thought the exercise demonstrations were a good way to teach how to complete the exercises. When asked if there were any benefits to the guided exercise program, several users mentioned seeing an improvement in their physical functioning, breathing, and motivation to exercise. Users also mentioned the convenience and flexibility to complete the PR exercises at their home/office at their own leisure was beneficial. Overall, users did not have problems using the video exercise instructions to complete their PR exercises. The most frequent suggestions to improve the exercise content were questions in between the exercises, decreasing the length of the exercise sessions, and including an option for a more challenging exercise program. The following citation demonstrates a patient perspective on how the interactive content facilitated patient daily exercise: *"Doing the exercises and the visual is better for me to focus on my breathing and how to do the exercises and to help me function a little better."*

3.2. Themes Related to the Interface of the Pulmonary Telerehabilitation System

Most users felt comfortable using the exercise system independently. A few users stated needing support from the research staff to troubleshoot the tablet when experiencing technical difficulties with the tablet (i.e., tablet freezing, internet connectivity issues, login issues). Users discussed several benefits of the home-based exercise system including the accessibility and convenience of being able to access the exercise system at home. Users also thought the PR program helped increase their motivation to exercise and created a daily purpose/goal. Two users responded that the feedback and online support of the physical therapist (PT) were valuable and helped improve their motivation. Nearly all users thought the introductory practice session administered prior to starting their personalized PR program was informative and sufficient. When asked about their preference for the presentation of exercise prompts and steps (audible, visual, or both), the majority of users preferred both the audio and visual prompts and steps.

Users were asked if the inclusion of more diverse content would make the experience (e.g., images, video clips) enjoyable. Many users agreed the addition of diverse content would make the experience more enjoyable, specifically if there were included with different exercises. Users compared the ease of use between the exercise program interface and their home TV remote control. Seven users responded that their TV remote control is easier to operate due to familiarity and ease of use. Five users stated that the exercise system is easier to navigate than their TV remote control. Three users thought that both the exercise system and their TV remote control are simple to operate. When users were asked if there was any aspect of the system that would make them refrain from using the exercise system to help with completing their exercises at home, users reported external factors such as health limitations and competing demands. Users also stated that limited motivation to exercise, difficulty finding an appropriate space to use the tablet, and connectivity issues were deterrents to using the exercise system. Users were asked how the exercise system can be improved. Three users suggested improving

functionality and connectivity of the tablet. Two users suggested including the option to change the level of difficulty and the length of the exercise sessions. The following citations exemplifies how the structured system interface facilitates patient engagement in pulmonary telerehabilitation: *"I think it is good because it gives me more structure about what to do, not just picking up a piece of paper and reading it. It gives me purpose"*

3.3. Themes Related to the Process of the Pulmonary Telerehabilitation System

Overall, users responded positively when asked how they felt about using a computer-guided program to complete the PR exercises. Users stated that they like the idea of using a computer-guided individualized exercise program at home. When compared to going to the gym or using other exercise programs/tools, 11 users thought the exercise program was better because it was home-based, cost-effective, decreased the risk of COVID-19, and allowed users to review the exercise demonstrations at their own pace. One user stated that the two-way remote communication with the telerehab team and feedback from PT helped improve their motivation. Majority of users thought that the PR program will help improve their physical ability to exercise, motivation to complete the exercises, and maintain a workout routine. Nearly all users reported that the exercise program would improve their confidence in completing the exercises at home.

When users were asked to describe what the best feature exercise program is, users responded that the increase in motivation to exercise and the convenience and flexibility of the at-home exercise program were beneficial. One user stated that submitting an exercise log and being monitored by the PT was beneficial. Most users reported having no problems or concerns with using the exercise program. Three users mentioned experiencing issues with the tablet including the transmission of data and connectivity.

Users were asked what additional features could be added to the exercise program to make the system more effective. Most users reported no suggestions to make the exercise program more effective. Four users suggested increasing the variety of exercises, additional breathing exercises, an option to increase the level of exercise difficulty, and making the system more interactive. Two users mentioned including the option to change the background music. The following citation demonstrates patient perspective on the telerehabilitation process enabling ongoing participation in the program: *"The major benefit is that you feel that there is someone there helping you. I felt that there was a person there helping me."*

4. Discussion

Past studies investigating patient perspectives on home-based PR consisted of an 8-week intervention. In this qualitative study, we examined the perceptions of patients regarding their participation in a 12-month home-based pulmonary telerehabilitation program.
The salient features of the exercise program that emerged from the qualitative data related to the accessibility of the guided exercise instructional videos, the ability to review the exercise demonstrations and self-correct, the convenience of the at-home exercise program, and perceived improvement in physical health, breathing, and motivation to exercise. Patients felt comfortable using the exercise system to complete their prescribed exercises. Despite the reported concerns with tablet operation, system connectivity, and setbacks brought on by health restrictions, patients persisted in making an attempt to overcome any technical issues or seek adjustments to their workout regimen

as necessary. The consistent feedback and communication with the PT helped improve the motivation to exercise and continue participation in the home-based program with minimal disruptions. Overall, COPD patients have enthusiastically embraced the telerehabilitation program, rating it highly for usefulness and satisfaction. Many patients expressed the desire to continue using the exercise system beyond the 12-month study period.

5. Conclusion

We evaluated an online pulmonary telerehabilitation platform with 15 COPD patients. The findings of this qualitative study show that patients had a positive acceptance and high levels of satisfaction with the pulmonary telerehabilitation system which allowed them to successfully follow their individualized exercise programs for 12 months. This study offers a thorough investigation of patient viewpoints when utilizing telerehabilitation technology. These findings will be considered during the future design and deployment of a patient-centered COPD telerehabilitation system tailored to patients' requirements and preferences, and it can be extended to other rehabilitation areas [9-10].

Acknowledgements

The project was funded by NIH grant R33HL143317.

References

[1] Mei F, Dalmartello M, Bonifazi M, Bertuccio P, Levi F, Boffetta P, Negri E, La Vecchia C, Malvezzi M. Chronic Obstructive Pulmonary Disease (COPD) mortality trends worldwide: An update to 2019. Respirology. 2022;2(11):941-50.

[2] Zeng X, Chen H, Ruan H, Ye X, Li J, Hong C. Effectiveness and safety of exercise training and rehabilitation in pulmonary hypertension: a systematic review and meta-analysis. Journal of Thoracic Disease. 2020;12(5):2691.

[3] Lahham A, McDonald CF, Mahal A, Lee AL, Hill CJ, Burge AT, Cox NS, Moore R, Nicolson C, O'Halloran P, Gillies R. Home-based pulmonary rehabilitation for people with COPD: a qualitative study reporting the patient perspective. Chronic respiratory disease. 2018;15(2):123-30.

[4] Holland AE, Mahal A, Hill CJ, Lee AL, Burge AT, Cox NS, Moore R, Nicolson C, O'Halloran P, Lahham A, Gillies R. Home-based rehabilitation for COPD using minimal resources: a randomised, controlled equivalence trial. Thorax. 2017 ;72(1):57-65.

[5] Weller SC, Vickers B, Bernard HR, Blackburn AM, Borgatti S, Gravlee CC, Johnson JC. Open-ended interview questions and saturation. PloS one. 2018 Jun 20;13(6):e0198606.

[6] Bedra M, McNabney M, Stiassny D, et al. Defining patient-centered characteristics of a telerehabilitation system for patients with COPD. Stud Health Technol Inform. 2013;190:24-6.

[7] Braun V, Clarke V. Using thematic analysis in psychology. Qualitative research in psychology. 2006;3(2):77-101.

[8] Ritchie J. and Spencer L. Qualitative data analysis for applied policy research. Analyzing Qualitative Data. 1994; 9:1-22.

[9] Feasibility of post-acute hip fracture telerehabilitation in older adults. Stud Health Technol Inform. 2015;210:469-73.

[10] Jeong IC, Karpatkin H, Finkelstein J. Physical Telerehabilitation Improves Quality of Life in Patients with Multiple Sclerosis. Stud Health Technol Inform. 2021;284:384-388.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230323

Telemedicine Based on Human Activity Recognition in Elderly Healthcare

Raheleh GHOUCHAN NEZHAD NOOR NIA[a], Afrooz ARZEHGAR[a], Vajiheh DEHDELEH[a] and Saeid ESLAMI[a,1]
[a]*Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
ORCiD ID: Raheleh Ghouchan Nezhad Noor Nia https://orcid.org/0000-0001-5697-9041, Afrooz Arzehgar https://orcid.org/0000-0003-0459-3395, Vajiheh Dehdeleh https://orcid.org/0000-0003-2767-7321, Saeid Eslami https://orcid.org/0000-0003-3755-1212

Abstract. Nowadays, telemedicine can provide remote clinical services for the elderly, using smart devices like embedded sensors, via real-time communication with the healthcare provider. In particular, inertial measurement sensors such as accelerometers embedded in smartphones can provide sensory data fusion for human activities. Thus, the technology of Human Activity Recognition can be applied to handle such data. In recent studies, the three-dimensional axis has been used to detect human activities. Since most changes in individual activities occur in the x- and y-axis, the label of each activity is determined using a new two-dimensional Hidden Markov Mode based on these two axes. To evaluate the proposed method, we use the WISDM dataset which is based on an accelerometer. The proposed strategy is compared to General Model and User-Adaptive Model. The results indicate that the proposed model is more accurate than the others.

Keywords. Human Activity, Healthcare, Bayesian Networks, Markov Model, Recognition

1. Introduction

Medical and assistive systems utilizing wearable sensors are being employed to provide long-term care and enhance the quality of life for the elderly [1]. Human Activity Recognition (HAR) based on wearable sensors, such as those found in smartphones, is being used to monitor patients in hospitals and at home. However, the most significant challenges faced by HAR are scalability, complex actions, and human behaviors in a complex environment [2]. Researchers have proposed several approaches for activity recognition, including Ameva, which uses selection, discretization, and classification techniques [3]. User-adaptive models (UAM) utilize deep transfer learning and data augmentation to improve prediction performance with limited training data [4]. Furthermore, a new method has been presented that integrates personal experience into the HMM for activity recognition using a personal wearable computer eButton, which includes multiple sensors [5]. Additionally, some approaches for HAR using deep

[1] Corresponding Author: Saeid Eslami, PhD., Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; E-mail: EslamiS@mums.ac.ir.

learning techniques have been introduced. In [6], a model is presented that combines residual block and bi-directional LSTM to extract features from multidimensional signals of microelectromechanical system inertial sensors, and the features are classified using a Softmax layer. In [7], a method based on Encoder-Decoder Convolutional Neural Networks is introduced, which uses both publicly labeled and private unlabeled raw sensor data to extract relevant features and enhance knowledge in the innermost layers. In [8] an ensemble approach called EnsemConvNet is presented, which combines three classification models to predict human activity based on time series data. Finally, in [9], a new approach is introduced that uses inertial wearable sensors to detect and classify human activities by jointly segmenting multidimensional time series using an HMM in a multiple regression context. The proposed method in this study aims to use smartphone-based accelerometers to detect physical human activities and categorize similar activities using Bayesian Networks (BNs). Additionally, a two-dimensional HMM is used to determine the label of each activity based on the x- and y-axis. The rest of this study is organized as follows. The proposed approach is explained in Section 2. The experimental results present in Section 3. Section 4 concludes and describes the ongoing work to improve the proposed approach.

2. Proposed Method

In this study, a set of motion data received from phone-based accelerometers is used to identify the user's physical activity. People were asked to walk, run, climb stairs, descend stairs, sit, and stand while holding their cell phone. It is clear that sitting and standing do not exhibit distinct behavior pattern while the other four activities include repetitive movements and show periodic behavior [9]. In this study, detecting different physical activities by analyzing the patterns in the accelerometer data of participants performing six activities for four activities but not for sitting and standing is analyzed.

The proposed strategy has two phases, in the first phase, we use BNs to classify time series data into three groups based on some features [10]. A BN is formally shown by a pair $B = \langle G, \Theta \rangle$ for the set U so that G represents a directed acyclic graph whose vertices are the random variables $X_1, \dots, X_n$ and edges in this network represent dependencies between these variables [11]. The second component Θ is related to parameters which quantify the network. For each possible value x_i, this parameter contains $\theta_{x_i|\Pi_{x_i}} = P_B(x_i|\Pi_{x_i})$, and Π_{x_i} of Π_{X_i}, where Π_{X_i} represents the parents of X_i. A unique joint probability distribution over U is given by (11) as follow.

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^{n} P_B\left(X_i|\Pi_{X_i}\right) = \prod_{i=1}^{n} \theta_{X_i|\Pi_{X_i}} \quad (1)$$

In this paper we use BNs to identify possible relationships to evaluate the membership class. In this strategy, the group of "Slow running and walking" belongs to same category, "climbing and descending stair" is classified in another class, and "sitting and standing" is located in the last group. In the second phase, the HMM is designed and trained for each data class. The formal definition of an HMM is as follows [12]:

$$\lambda = (A, B, \pi) \quad (2)$$

A is a transition array, storing the probability of state j following state i. Note the state transition probabilities are independent of time (12):

$$A = \left[a_{ij}\right], a_{ij} = P(q_t = s_j|q_{t-1} = s_i) \quad (3)$$

B is the observation array, storing the probability of observation k being produced from the state j, independent of t (12):

$$B = [b_i(k)], b_i(k) = P(x_t = v_k | q_t = s_i) \quad (4)$$

π is the initial probability array (12):

$$\pi = [\pi_i], \pi_i = P(q_1 = s_i) \quad (5)$$

Since a person's movement is mostly observable in the x and y axes, in this paper we use two-dimensional HMM with time series information of these two axes to detect the label of each data. The speed of time series is used as an observation to predict the label. The process of the proposed method is shown in Figure 1.

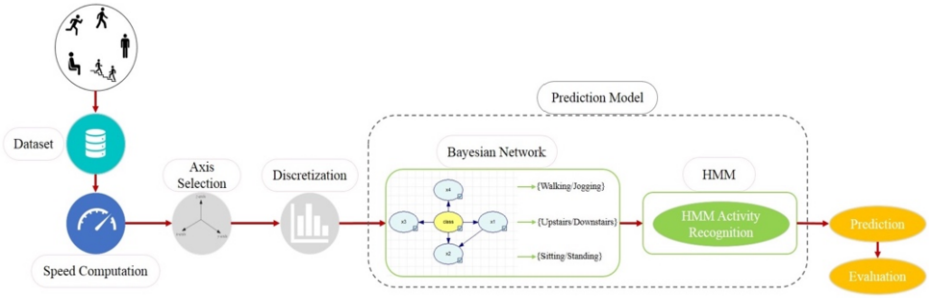


Figure 1. The process of the proposed method

3. Experimental results

To evaluate the performance of the proposed strategy, we use WISDM dataset of 36 users. In this dataset the information about each person and its label recorded [10]. Before collecting dataset, the ethical approval was obtained from the Fordham University IRB (Institutional Review Board) since the study involved "experimenting" on human subjects and there was some risk of harm [10]. The dataset contains six classes of walking, Jogging, Upstairs, Downstairs, Sitting and Standing which are presented by class 1 to 6, respectively. The activity changes of each class in three axes x, y and z are available. In the phase of designing, 70% of the data is considered for training, and the remaining 30% is for testing. The results show that four features XABSOLDEV, YABSOLDEV, XSTANDDEV, and ZSTANDDEV have the most differences for recognizing the labels {1,2}, {3,4} and {5,6}; therefore, there is a good separation between the classes. Changes in XABSOLDEV feature are given in Figure 2, the changes of similar classes such as {1,2}, {3,4}, and {5,6} are in the same range, which indicates that similar activities are in the same category. Similar changes and interpretations are shown for YABSOLDEV, XSTANDDEV and ZSTANDDEV features in Figure 3, 4 and 5, respectively. In Figure 6, six classes for x-axis time series values are given. After detecting the class, the two-dimensional HMM is applied. Since a person's movement has the most changes in the x and y-axis, the two-dimensional Markov time series model with two features is used to investigate the effect of the two features at the same time. The result of evaluating the proposed method is given in Figure 7. The Area Under the Curve (AUC) is used to assess the performance of the model. The proposed method has been compared with the General Model (GM) [4] and User-Adaptive Model (UAM) [4] that the result of this comparison is shown in Table 1. According to Table 1, the proposed method is more efficient than two other models.

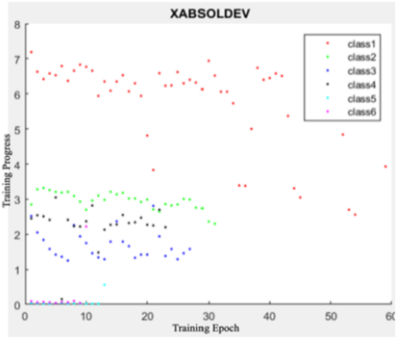


Figure 2. XABSOLDEV feature changes.

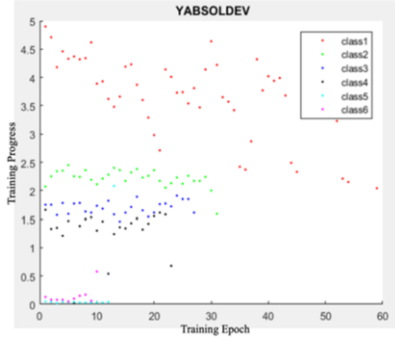


Figure 3. YABSOLDEV feature changes.

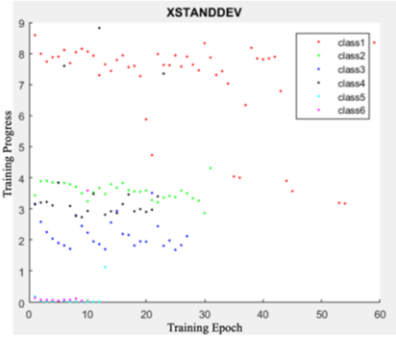


Figure 4. XSTANDDEV feature changes.

Figure 5. ZSTANDDEV feature changes.

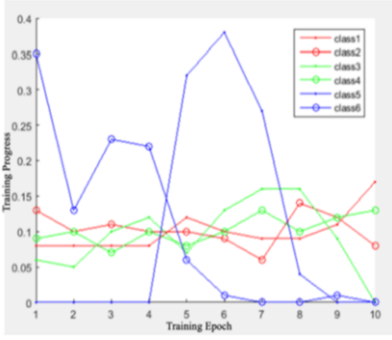


Figure 6. label prediction for each class.

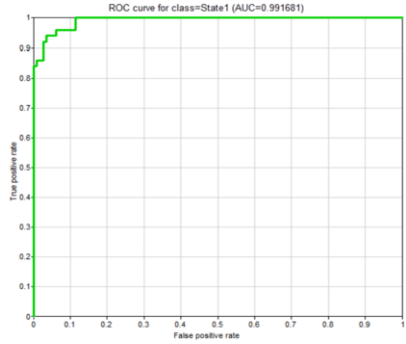


Figure 7. Proposed method evaluation.

Table 1. The accuracy of each class based on each activity for GM, UAM and the proposed method.

Class	**GM** (4)	**UAM** (4)	**Proposed method**
Walking	0.86	1.00	0.98
Jogging	0.75	0.83	0.95
Upstairs	0.28	0.67	0.93
Downstairs	0.29	0.64	0.91
Sitting	0.91	0.96	0.98
Standing	0.91	1.00	0.94
Accuracy	0.80	0.71	**0.94**

4. Conclusion

The aim of this paper is to present a model which be useful in monitoring the daily elderly activities in HAR system. In this study, two-stage classification is used, at first, the model categorizes similar activities by BN into three groups based on the features. In this model, the category of "Slow running and walking" belongs to same category, "climbing and descending stair" is classified in another class, and "sitting and standing" is located in the last group. Then to detect the label of each data, two-dimensional HMM with two features with time series information of the x and y axes are used to investigate the effect of these two features at the same time because a person's movement is mostly observable in these two axes. To evaluate the proposed strategy in this paper, it is compared to GM and UMA methods. The results of the tests indicate that the new model with 94% accuracy is more accurate than other methods.

References

[1] United Nations D of E. WHO Global Report on Falls Prevention in Older Age. Community Health [Internet]. 2007 [cited 2023 Feb 22];53. Available from: https://books.google.com/books/about/WHO_Global_Report_on_Falls_Prevention_in.html?id=ms9o2dvfaQkC

[2] Serpush F, Menhaj MB, Masoumi B, Karasfi B. Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System. Comput Intell Neurosci. 2022;2022.

[3] Álvarez de la Concepción MÁ, Soria Morillo LM, Álvarez García JA, González-Abril L. Mobile activity recognition and fall detection system for elderly people using Ameva algorithm. Pervasive Mob Comput. 2017 Jan 1;34:3–13.

[4] Garcia-Ceja E, Riegler M, Kvernberg AK, Torresen J. User-adaptive models for activity and emotion recognition using deep transfer learning and data augmentation. User Model User-adapt Interact [Internet]. 2020 Jul 1 [cited 2023 Feb 22];30(3):365–93. Available from: https://link.springer.com/article/10.1007/s11257-019-09248-1

[5] Li Z, Wei Z, Yue Y, Wang H, Jia W, Burke LE, et al. An Adaptive Hidden Markov Model for Activity Recognition Based on a Wearable Multi-Sensor Device. J Med Syst [Internet]. 2015 May 1 [cited 2023 Feb 22];39(5):1–10. Available from: https://link.springer.com/article/10.1007/s10916-015-0239-x

[6] Li Y, Wang L. Human Activity Recognition Based on Residual Network and BiLSTM. Sensors 2022, Vol 22, Page 635 [Internet]. 2022 Jan 14 [cited 2023 Feb 22];22(2):635. Available from: https://www.mdpi.com/1424-8220/22/2/635/htm

[7] Hurtado S, García-Nieto J, Popov A, Delgado IN, Navas-Delgado I. Human Activity Recognition From Sensorised Patient´s Data in Healthcare: A Streaming Deep Learning-Based Approach. Article in International Journal of Interactive Multimedia and Artificial Intelligence [Internet]. 2022 [cited 2023 Feb 22]; Available from: http://dx.doi.org/10.9781/ijimai.2022.05.004

[8] Mukherjee D, Mondal R, Singh PK, Sarkar R, Bhattacharjee D. EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. Multimed Tools Appl [Internet]. 2020 Nov 1 [cited 2023 Feb 22];79(41–42):31663–90. Available from: https://link.springer.com/article/10.1007/s11042-020-09537-7

[9] Trabelsi D, Mohammed S, Chamroukhi F, Oukhellou L, Amirat Y. An unsupervised approach for automatic activity recognition based on Hidden Markov Model regression. IEEE Transactions on Automation Science and Engineering. 2013 Jul;10(3):829–35.

[10] Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations Newsletter [Internet]. 2011 Mar 31 [cited 2023 Feb 22];12(2):74–82. Available from: https://dl.acm.org/doi/10.1145/1964897.1964918

[11] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. Mach Learn [Internet]. 1997 [cited 2023 Feb 22];29(2–3):131–63. Available from: https://link.springer.com/article/10.1023/A:1007465528199

[12] Blunsom P. Hidden Markov Models. 2004.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230324

Interpretable EEG-Based Emotion Recognition Using Fuzzy Cognitive Maps

Georgia SOVATZIDI[a] and Dimitris K. IAKOVIDIS[a,1]
[a] *Dept of Computer Science and Biomedical Informatics, Univ. of Thessaly, Greece*

Abstract. The brain is one of the most complex parts of the human body, consisting of billions of neurons and it is involved in almost all vital functions. To study the brain functionality, Electroencephalography (EEG) is used to record the electrical activity generated by the brain through electrodes placed on the scalp surface. In this paper, an auto-constructed Fuzzy Cognitive Map (FCM) model is used for interpretable emotion recognition, based on EEG signals. The introduced model constitutes the first FCM that automatically detects the cause-and-effects relations existing among brain regions and emotions induced by movies watched by volunteers. In addition, it is simple to implement and earns the trust of the user, while providing interpretable results. The effectiveness of the model over other baseline and state-of-the-art methods is examined using a publicly available dataset.

Keywords. Electroencephalography (EEG), Emotion Recognition, Fuzzy Cognitive Map (FCM), Fuzzy Logic, Interpretability

1. Introduction

Emotions are associated with many tasks in human cognition, including decision-making processes, communication, perception, and intelligence. To detect emotions, functional neuroimaging techniques, such as EEG, are used to study the electrical activity of the brain of humans, while receiving stimuli, e.g., film clips, music, pictures [1]. Machine learning techniques have been widely used to deal with the challenging task of emotion recognition [2]. However, such approaches are considered as "black boxes", as they do not provide sufficient explanations for the resulting outcome.

During the last decades, the use of fuzzy network structures has shown great potential in various fields, including medicine [3]. Specifically, FCMs are graph-based methods, which have received considerable attention from researchers, due to their simplicity, effectiveness, and high ability to deal with uncertainties [4]. However, a limitation of FCMs is that there is a need for human participation to determine the structure of the graph. Recent modified FCMs have been effectively applied to various medical problems, including Constructive Fuzzy Representation Model (CFRM) for heart disease classification [5], and Constructive FCM (CFCM) for depression severity estimation [4]. In this paper, we introduce an FCM model for interpretable emotion recognition, based on EEG signals. Unlike conventional FCMs, the proposed model contributes to automatically detect the cause-and-effects relations that exist between its concepts without the need for human intervention, from the datasets used in each

[1] Corresponding Author: Dimitris K. IAKOVIDIS, E-mail: diakovidis@uth.gr.

experiment. Regarding the emotion recognition task, the proposed FCM model has the following contributions: a) it automatically detects the relations existing among the examined brain regions and the emotions induced by movies watched by volunteers; b) it is simple to implement; and, c) it provides understandable results.

2. Materials and Methodology

2.1. Dataset

To perform emotion recognition based on the proposed FCM, the DREAMER dataset was used [1]. It is a multimodal dataset consisting of EEG and electrocardiography (ECG) signals obtained from 23 volunteers, while watching 18 film clips selected to elicit certain emotions. For our experiments, we utilize the EEG recordings (128 Hz) of the dataset, which were collected from 14 EEG channels and analyzed in terms of arousal, and valence [1]. Depending on the evoked valence and arousal values, the 2-D valence-arousal space is derived (Figure 1). Each quadrant represents one of the following possible combinations of High (H)/Low (L) Valence (V)/Arousal (A) states and includes relative emotions. For example, as can be observed from Figure 1, HAHV includes positive emotions, *e.g.,* happiness and amusement. In addition, to examine the emotional state of the brain, a segmentation of the EEG electrode positions into brain regions was performed, based on [4], and adapted to the dataset used.

2.2. Dataset

To conduct the experiments, the EEG signals were preprocessed, based on [1]. Specifically, the signals were filtered between 4 and 48 Hz, using a FIR filter with a Hamming window of 212 samples. Moreover, an artefact rejection process was performed, *i.e.*, Artefact Subspace Reconstruction (ASR) [1], utilizing the EEGLab toolbox [6]. Then, the EEG signals were separated into the following frequency bands: theta (4 Hz - 8 Hz), alpha (8 Hz - 13 Hz), and beta (13 Hz - 20 Hz). In addition, the Power Spectral Densities (PSDs) of the EEG signals in different frequency bands were calculated, given that they are significantly correlated with human emotions [1]. All the extracted features are concatenated into a final feature vector $F_{r,\lambda}^{w} = \left(f_1^w, f_2^w, \dots, f_{N_r}^w\right)$, where w corresponds to the frequency band, $r = 1, \dots, R$ are the examined segmented brain regions, N_R is the number of EEG electrodes in each r, and $\lambda = 1,2, \dots, \Lambda$ represents the class of the examined problem. In our experiments, λ corresponds to positive and negative emotion. The fuzzy set construction is performed, aiming to characterize linguistically the calculated brain electrical activity, based on [4]. Specifically, a clustering algorithm is applied to group $F_{r,\lambda}^{w}$ into a set of M clusters with $M < V_\lambda$, where V_λ corresponds to the total number of volunteers participating in the emotion recognition problem. The resulting centroids $q_m, m = 1, \dots, M$ are sorted ascendingly and their standard deviations $\sigma_m, m = 1, \dots, M$ are used to define the fuzzy sets $\Phi_r^m, m = 1, \dots, M, r = 1, \dots, R$. These fuzzy sets have triangular membership functions, where the top of the triangle is located at the q_m, and the base is extended to the range $[q_m - \sigma_m, q_m + \sigma_m]$. Each fuzzy set corresponds to a linguistic term, *e.g.*, for $M = 3$ the linguistic terms are "*Low*", "*Medium*", "*High*", which is described by a membership function $\mu_r^m(f_{v,r}^w)$, v =1,2,…, V_λ.

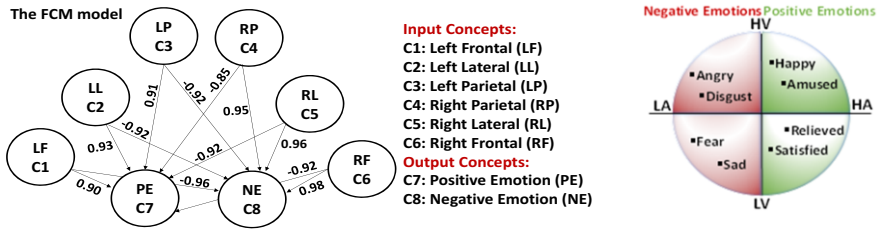


Figure 1. Proposed FCM structure (left). 2D valence-arousal emotion space -High (H) / Low (L) Arousal (A), Valence (V) (right).

2.3. FCM Construction

An FCM is a graph-based model consisted of concept nodes, $C_i = C_1, C_2, \ldots, C_K$, where K is the total number of concepts, and weighted arcs w_{ij} that determine the interconnection between C_i to C_j. The concept values of nodes represent the state vector $A^t = (A_1^t, \ldots, A_K^t)$. Thus, regarding the proposed FCM structure, the input concepts (C_1-C_6) of the graph represent the brain regions, *i.e.*, Left (L)/ Right (R) Frontal (F), Parietal (P), Lateral (L), as depicted in Figure 1 (left). The output concepts (C_7, C_8) represent the Positive and Negative Emotion (PE/ NE), which is correlated with the corresponding arousal/valence values (Figure 1). The interconnections between two concepts C_i and C_j , $i \neq j$, are defined in relation to the differences observed in brain activity in the examined brain regions, regarding positive and negative emotions. Specifically, the weight of each edge $i \rightarrow j$ is calculated based on the membership functions obtained by the respective fuzzy sets Φ_r^m. The fuzzy sets are then aggregated, using a union operation and the calculation of their center of gravity $\left(g_{i,j}\right)$ follows. For example, for the calculation of w_{16}, we calculate the center of gravity of fuzzy sets that correspond to C_1 and C_6 ,*i.e.,* Φ_1^m and Φ_6^m . The calculation of interconnections between two input concepts C_i and C_j , $i \neq j$ is given by: $w'_{ij} = \frac{max_{m=1}^{M}\left(\mu_i^m(g_{i,j}),\mu_j^m(g_{i,j}\right)}{\arg\left(max_{m=1}^{M}\left(\mu_i^m(g_{i,j}),\mu_j^m(g_{i,j})\right)\right)}$. Similarly, the calculation of interconnection between an input and an output concept is given by $w_{ij} = \frac{max_{m=1}^{M}\left(\mu_i^m(g_{i,j})\right)}{\arg\left(max_{m=1}^{M}\left(\mu_i^m(g_{i,j})\right)\right)}$. In the constructed graph of the examined problem (Figure 1), all the input concepts have interconnections among them. However, for illustrative purposes, the interconnections among the input and output concepts have been selected for demonstration. The proposed FCM model iteratively calculates its states until convergence, based on $A_i^{t+1} = h\left(A_i^t + \sum_{j=1,j\neq i}^{n} w_{ij}A_j^t\right)$, where $t = 1 \ldots., T$ is the iteration number, w_{ji} is the weight matrix of C_i to C_j, and h is a transfer function. The values of the initial state vector $A^0 = (A_1^0, \ldots, A_K^0)$ are estimated based on [4].

3. Results and Discussion

3.1. Classification Outcomes

The proposed model was evaluated in terms of its decision-making performance with

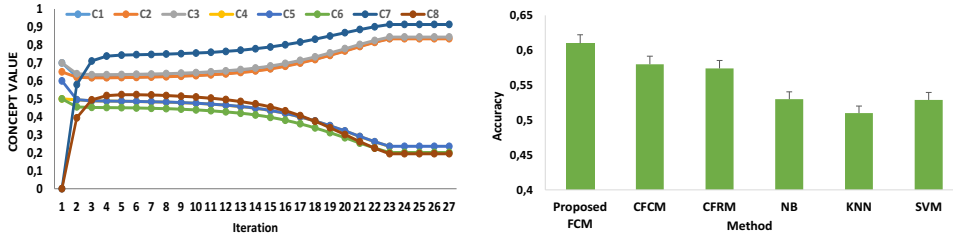


Figure 2. (left) Convergence plot of the proposed FCM model. (right) Comparisons of the average accuracy and standard deviation using the PSD features.

well-known classifiers, *i.e.,* Naïve Bayes (NB), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), as well as state-of-the-art approaches, *i.e.,* CFCM , CFRM [3], [4]. As it can be observed from Figure 2 (right), the proposed model outperforms the rest methods in terms of average accuracy, using the calculated PSD features, while providing interpretable results. In addition, it reaches a steady state and finally converges, after 23 iterations. The convergence plot is depicted in Figure 2 (left).

3.2. Classification Outcomes

To better understand the interpretable emotion recognition based on the proposed FCM model, the following indicative example is included in this section. Let us consider a randomly selected volunteer from the dataset, for automatically detecting and interpreting his emotional state. The initial state vector $A^0 = (0.70, 0.65, 0.60, 0.60, 0.60, 0.50, 0, 0\ )$ is calculated as described in the methodology and is inserted into the FCM to start the reasoning process. After $k = 23$ iterations, the FCM converges into a steady state, which results in $A^{23} = (0.86, 0\ 85, 0.86, 0.21, 0.21, 0.18, 0.92, 0.17)$. Specifically, regarding A^0 and A^{23}, the first three numbers correspond to the concept values of C_1, C_2, C_3, which represent the Left Hemisphere of the Brain, *i.e.*, Left Frontal Lateral, and Temporal regions, while C_4, C_5, C_6 represent the Right Hemisphere, respectively (Figure 1). The last two values of A^0 and A^{23} represent the output concept values, *i.e.*, the values of C_7 and C_8 that correspond to the positive and negative emotions (Figure 1). To interpret the outcome in a way compatible to human logic, the calculated PSDs are then characterized linguistically, using fuzzy sets. Specifically, the PSDs are described using the following four linguistic terms: "*Low*" (L)=[0, 0.4], "*Medium*"(M)=[0.2, 0.6], "*High*" (H)=[0.4, 0.8], "*Very High*" (VH)=[0.6, 1]. Consequently, looking at A^{23} the examined volunteer has a total higher electrical activity in the left-brain regions compared to the right ones. Specifically, Left Frontal Lateral and Parietal Regions resulted in "*Very High*" PSDs, with corresponding calculated concept values $C_1 = 0.86, C_2 = 0.85, C_3 = 0.86$. Regarding the Right Frontal Lateral and Parietal Regions, they are characterized by "*Medium*" PSDs, as the respective concept values are $C_4 = 0.21, C_5 = 0.21, C_6 = 0.18$. The values of the outcome concepts, which determine the final results, are $C_7 = 0.92$ and $C_8 = 0.17$, which is linguistically described as "*Very High*" Positive Emotion, "Low" Negative Emotion. An interpretation of the proposed FCM model (Figure 1 (left)) considering the calculated interconnections, regarding the examined case, follows:

- Left-brain regions have positive relations ($w > 0$) with the positive feelings, such as joy or happiness, and negative ($w < 0$) with the negative feelings, *e.g.,* fear or disgust. Consequently, an increase in the electric activity of the left-brain region

Figure 3. Scalp maps that show the scalp distribution of power at 6 Hz on the left-brain region (left), and right brain region (right).

evokes an increase (decrease) in the degree of the positive (negative) emotional state.

- Right-brain regions have positive relations ($w > 0$) with the negative emotional states, and negative ($w < 0$) with the positive feelings. Thus, an increase in the activity of the right-brain region evokes an increase (decrease) in the degree of the negative (positive) emotional state.

Figure 3 illustrates the scalp distribution of power, at 6 Hz. The warm (cold) color in scalp topography indicates high (low) energy, respectively.

4. Conclusions

In this paper, an auto-constructed FCM model was proposed for interpretable emotion recognition, based on EEG signals. The presented graph-based model automatically detects the cause-and-effects relations between the examined brain regions and the emotions induced to volunteers, after having watched movie clips. In addition, it is simple to implement and earns the trust of the user, while providing interpretable results. Future work includes further investigation of the proposed framework on various domains, using different types of membership functions and datasets.

Acknowledgment: We acknowledge support of this work by the project “Smart Tourist” (MIS 5047243) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Program "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

References

[1] Katsigiannis S, Ramzan N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE journal of biomedical and health informatics. 2017 Mar 27;22(1):98-107.

[2] Zhang J, Yin Z, Chen P, Nichele S. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion. 2020 Jul 1;59:103-26.

[3] Vlamou E, Papadopoulos B. Fuzzy logic systems and medical applications. AIMS neuroscience. 2019;6(4):266.

[4] Sovatzidi G, Vasilakakis M, Iakovidis DK. Constructive Fuzzy Cognitive Map for Depression Severity Estimation. Studies in health technology and informatics. 2022 May 25;294:485-9.

[5] Vasilakakis MD, Iakovidis DK, Koulaouzidis G. A Constructive Fuzzy Representation Model for Heart Data Classification. InMIE 2021 May 1 (pp. 13-17).

[6] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods. 2004 Mar 15;134(1):9-21.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230325

Client-Side Application of Deep Learning Models Through Teleradiology

Sébastien JODOGNE[1]
ICTEAM, UCLouvain, Belgium
ORCiD ID: https://orcid.org/0000-0001-6685-7398

Abstract. Deep learning models for radiology are typically deployed either through cloud-based platforms, through on-premises infrastructures, or though heavyweight viewers. This tends to restrict the audience of deep learning models to radiologists working in state-of-the-art hospitals, which raises concerns about the democratization of deep learning for medical imaging, most notably in the context of research and education. We show that complex deep learning models can be applied directly inside Web browsers, without resorting to any external computation infrastructure, and we release our code as free and open-source software. This opens the path to the use of teleradiology solutions as an effective way to distribute, teach, and evaluate deep learning architectures.

Keywords. Medical imaging, teleradiology, deep learning, WebAssembly

1. Introduction

Deep learning applied to radiology has attracted a lot of interest in the recent years [1]. Convolutional neural networks (CNN) have proved to be an extremely promising building block for the tasks of classification, segmentation, localization, and detection applied to medical images [2]. In particular, the U-Net architecture is considered as a state-of-the art CNN for 2D biomedical image segmentation [3], with extensions to dense 3D segmentation [4]. However, the deployment of successful deep learning models in clinical setups involves multiple challenges that are notably related to the software distribution of the models, to the security of clinical data, and to the certification and reimbursement of associated clinical decision support systems.

Nowadays, the most encountered type of deployment for deep learning models is cloud-based. In this approach, hospitals internally setup a gateway that implements on-demand auto-routing of DICOM images from their PACS (Picture Archiving and Communication System) to a proprietary cloud platform managed by the vendor. After the analysis of the images, the cloud platform typically sends back a DICOM series that contains a PDF report, possibly associated with structured reports or annotated secondary capture images, which can be reintegrated into the PACS. The DICOM gateway is often implemented on the top of the DICOMweb protocol over HTTPS communications to protect the confidentiality of patient data. Despite its popularity, this approach can be

[1] Corresponding Author: Sébastien Jodogne Computer Science and Engineering Department (INGI), Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium, E-mail: sebastien.jodogne@uclouvain.be.

slow because of the time that is necessary to transmit the hundreds of megabytes of a single DICOM study over Internet, and the delay for getting the results of the analysis can largely vary depending upon the workload of the cloud infrastructure, which might be problematic depending on the clinical application. Furthermore, using cloud platforms implies a centralization of imaging studies, which necessitates complex legal agreements with the hospital, and a careful de-identification to avoid leaks of personal data [5]. Finally, as the usage of deep learning increases, hospitals want to take advantage of multiple deep learning algorithms, which causes scalability issues that could be mitigated through the deployment of shared gateways communicating with multiple vendors.

An alternative to cloud-based platforms consists in deploying the deep learning models as central, on-premises software running directly inside the computing infrastructure of the hospital. This approach has the major advantage of restraining the network exchanges to the intranet of the hospital, which guarantees higher network bandwidth, reserved computational power, and security by design in the handling of patient data. On-premises deep learning algorithms can be distributed in two distinct ways: either directly integrated within the PACS, or maintained as a separate solution by a third-party vendor. The former distribution method ensures a single point of contact for the hospital, but is not proposed by all the PACS vendors, and provides a restricted portfolio of algorithms selected by the vendor. The latter method opens a larger choice of algorithms, but requires more complex and costly integrations, as each third-party vendor must install and maintain its own dedicated computing infrastructure inside the hospital.

A third possibility consists in shipping the deep learning models together with the heavyweight viewers installed on imaging workstations, making local computational resources profitable. This avoids the installation of the dedicated computer clusters, which contrasts with on-premises infrastructures, while keeping the DICOM instances inside the intranet. This solution, however, vastly complicates the deployment of the algorithms, as it requires the local team of system administrators to keep the software and the models continuously up-to-date on the target computers. Furthermore, for security reasons, it should stay limited to fixed workstations that are directly connected to the PACS through DICOM query/retrieve, which is typically not the case with physicians' laptops. This type of deployment is also still dependent on the portfolio of deep learning models that are provided by the vendors of heavyweight viewers.

Besides their complexity and cost, the three types of deployments described above are inherently aimed at radiologists working inside state-of-the-art hospitals. This raises concerns about the need of democratizing deep learning models by sharing the technical knowledge behind such algorithms to a more general audience. Indeed, despite a general call to use deep learning, few skilled workers have the opportunity to experiment with machine learning algorithms during their training or as a part of their continuing education. Such technologies are also hardly affordable for emerging economies. Furthermore, even if there exists free and open-source software for medical imaging with advanced deep learning features [6], physicians, or more generally healthcare professionals, expect simple, user-friendly interfaces that are entirely focused on the application of models. It also remains challenging to evaluate the relevance of deep learning models for healthcare, because of the gap between research software and clinical environments. According to this discussion, this paper introduces a free and open-source teleradiology solution that can be used to distribute, then apply deep learning models directly inside Web browsers, without the need of deploying a dedicated computation infrastructure.

2. Methods

Deep learning is inherently asymmetrical: The training of models requires much data and computational power, notably provided by graphics processing units (GPU), whereas their application is within the reach of any modern computer. This allows to envision the application of deep learning architectures by Web browsers, if care is taken to optimize the operations on tensors wrt. a straight JavaScript implementation.

To this end, we propose to rely on WebAssembly, which is an emerging Web technology that has been designed for situations in which complex computations must be carried on by a Web browser. WebAssembly is an open standard specified by W3C that defines a portable binary code (bytecode) to create high-performance Web applications. Thanks to the Emscripten compiler, C++ source code can be converted to WebAssembly bytecode, which is then executed at almost native speed, i.e., faster than JavaScript, by a stack virtual machine running inside the Web browser. WebAssembly is actively backed by the industry and is nowadays supported by all the major Web browsers. It has recently been demonstrated that WebAssembly can be used to develop a cross-platform C++ library for the rendering of DICOM images that is compatible with desktop applications, mobile applications, as well as Web applications [7]. This novel library has been used to design the so-called "Stone Web viewer" as the first free and open-source implementation of a teleradiology solution leveraging WebAssembly.

The technical contribution of this paper consists in demonstrating that 2D U-Net models can be entirely applied within a teleradiology environment, in this case the Stone Web viewer. For this proof-of-concept, we have trained a sample U-Net model to segment the lungs on 2D images of a CT-scanner, using the images of the "*Lung CT Segmentation Challenge 2017*" (LCTSC) dataset [8]. This dataset contains DICOM CT-scan images together with the delineation of both lungs encoded as DICOM RT-STRUCT instances. The resulting U-Net model contains 7,759,521 floating-point parameters and was trained using the TensorFlow library [9] to minimize a smoothed version of the Dice loss function. The training process took 7 hours on a standard desktop computer (Intel Core i7-9700 CPU equipped with a NVIDIA GeForce GTX 1650 GPU). The resulting Dice score was 95%, which indicates average performance.

In order to run the segmentation process inside the Web browser, we have implemented a highly portable library written in plain C++ to apply CNN models. This library is fully compatible with WebAssembly (it is consequently single-threaded) and provides a reference implementation for all the distinct types of layers that are encountered in the U-Net architecture. Figure 1 explains the pipeline by which the models learned by TensorFlow are brought to the Stone Web viewer.

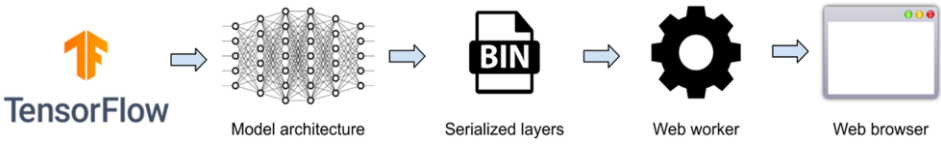


Figure 1. The pipeline used in this work. The U-Net model is first learned in Python using TensorFlow. A conversion script written in Python uses the TensorFlow API to explore the parameters and weights of the various layers in the trained model, then to serialize this information as a binary file using the Google's Protocol Buffers (Protobuf) cross-platform library. This preprocessing is done once and offline. Whenever the Stone Web viewer (that is executed by the Web browser) must apply the model to some image, it asks a HTML5 Web worker to load the model from the Protobuf file, to carry on the computations, then to generate an overlay from the result of the segmentation. This Web worker can be thought of as a "thread" that is executed in the background by the Web browser alongside the main user interface of the Stone Web viewer.

3. Results

The results of our work are depicted in Figure 2. As can be seen, we are able to segment CT-scan slices using our U-Net model, entirely inside the Web browser. The computation time is 11.3 seconds on a standard laptop computer equipped with an Intel i7-1165G7 CPU running Mozilla Firefox 109.0. Importantly, this computation only uses the CPU, so it is not required for the client computer to have dedicated local GPU hardware. Our C++ library for CNN uses the well-known "img2col" trick that rewrites convolutions as matrix products, hereby improving data locality [10]. If the direct "schoolbook" implementation of convolutional layers is used instead, the execution time slows down at 13.6 seconds. Our library also features a custom implementation of matrix product that leverages the optimized WebAssembly SIMD ("single instruction, multiple data"), since this set of instructions is currently not supported by standard C/C++ libraries for linear algebra. By comparison, executing the same code natively on the same laptop takes 5.4 seconds: This halving of the execution time is due to the fact that WebAssembly's SIMD instructions work on a width of 128 bits, whereas native AVX2 SIMD instructions work on a width of 256 bits.

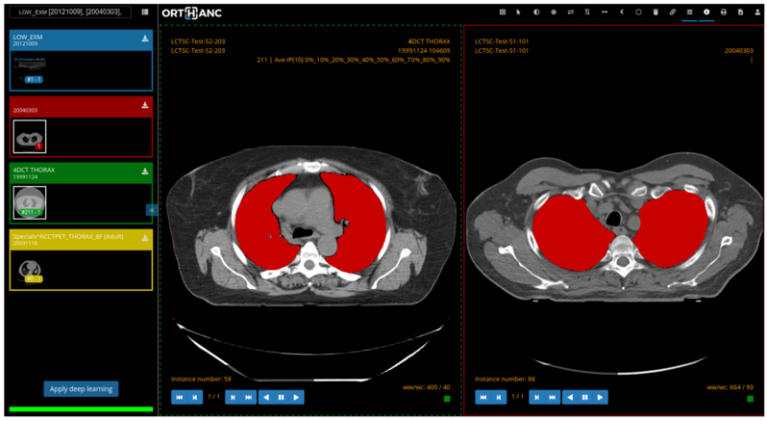


Figure 2. Segmentation using 2D U-Net models in the Stone Web viewer, applied to two test slices. The user launches the segmentation by clicking on the "*Apply deep learning*" button at the bottom of the Web interface. Then, a green progress bar below the button provides feedback about the computation by the Web worker. The segmentation mask is finally rendered as a red, semi-transparent overlay on the top of the slice.

4. Discussion

These results are promising and call for further work. For instance, this paper only considers the task of 2D segmentation using the U-Net architecture, but our library for deep learning in WebAssembly can be generalized to other tasks, such as classification, detection, or 3D segmentation. Further optimization of the CNN library is possible, by considering the specificity of the memory management of WebAssembly, by implementing more advanced techniques to compute convolutions [11], and by using WebGL to delegate parts of the computations to the GPU. Also, our CNN library could perfectly be embedded inside heavyweight viewers since it is written in plain C++.

Importantly, all the computations are done within the Web browser, so the DICOM instances never leave the PACS/viewer environment, making this deep learning solution

secure-by-design. This is notably a key advantage in the context of GDPR compliance that is implied by the MDR regulation. Furthermore, as the Stone Web viewer downloads its images using the DICOMweb standard, it is compatible with any modern PACS. Our work is released as free and open-source software in branch 3.0 of the Stone Web viewer. Interesting further research includes publishing a library of open-access models that could be used by our system, applying the system to selected clinical tasks, and evaluating the system for the training of skilled workers.

5. Conclusions

This paper demonstrates the feasibility of efficiently applying deep learning models client-side, directly by Web browsers, through a user-friendly teleradiology interface. This contribution allows to envision a scalable deployment of deep learning models, in the sense that these models could be immediately made available to many standard computers without installing any dedicated computation infrastructure or heavyweight software. It has also been shown that thanks to the WebAssembly SIMD, complex deep learning architectures such as U-Net with millions of parameters can be processed by a standard client computer in about ten seconds. This work is released as free and open-source software, in the hope that it will foster the evaluation artificial intelligence algorithms in clinical setups, and that it will promote the education of physicians, nurses, yet general audience, to artificial intelligence applied to medical imaging.

References

[1] Cheng PM, Montagnon E, Yamashita R, Pan I, Cadrin-Chênevert A, Perdigón Romero F, et al. Deep Learning: An Update for Radiologists. RadioGraphics. 2021;41(5):1427-45.
[2] Sarvamangala DR, Kulkarni RV. Convolutional Neural Networks in Medical Image Understanding: A Survey. Evolutionary Intelligence. 2022 jan;15(1):1-22.
[3] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, III WMW, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015) 18th International Conf. Munich, Germany, October 5 – 9, 2015, Proceedings, Part III. vol. 9351 of LNCS. Springer; 2015. p. 234-41.
[4] Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 4th International Conf. on 3D Vision (3DV). IEEE; 2016. p. 565-71.
[5] Paul M, Collberg C, Bambauer D. A Possible Solution for Privacy Preserving Cloud Data Storage. In: 2015 IEEE International Conference on Cloud Engineering. IEEE; 2015. p. 397-403.
[6] Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: Intraoperative Imaging and Image-Guided Therapy. Springer New York; 2013. p. 277-89.
[7] Jodogne S. Rendering Medical Images using WebAssembly. In: Proc. of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (Volume 2); 2022. p. 43-51.
[8] Yang J, Veeraraghavan H, Armato SG, Farahani K, Kirby JS, Kalpathy-Kramer J, et al. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. Medical Physics. 2018 Sep;45(10):4568-81.
[9] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al.. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015.
[10] Chellapilla K, Puri S, Simard P. High Performance Convolutional Neural Networks for Document Processing. In: Lorette G, editor. Tenth International Workshop on Frontiers in Handwriting Recognition. La Baule (France): Université de Rennes 1; 2006. p. 1-6.
[11] Georganas E, Avancha S, Banerjee K, Kalamkar D, Henry G, Pabst H, et al. Anatomy of High-Performance Deep Learning Convolutions on SIMD Architectures. In: SC18: International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE; 2018. p. 830-41.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230326

Technological Assessment of Smart Wearables and 5G-Integrated Edge Computing for Real-Time Health Monitoring

Paulo HAAS[a,1] and Thomas M. DESERNO[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hannover Medical School, Brunswick, Germany*

Abstract. Smart wearables advance to reliably and continuously measure vital signs. Analyzing the produced data requires complex algorithms, which would unreasonably increase the energy consumption of mobile devices and exceed their computing power. Fifth-generation (5G) mobile networks provide low latencies, high bandwidth, and many connected devices and introduced multi-access edge computing, which brings high computation power close to the clients. We propose an architecture for evaluating smart wearables in real-time and evaluate it exemplary with electrocardiography signals and binary classification of myocardial infarctions. Our solution shows that real-time infarct classification is feasible with 44 clients and secured transmissions. Future releases of 5G will increase real-time capability and enable capacity for more data.

Keywords. Health monitoring, 5G, edge computing, smart wearables

1. Introduction

Smart wearables can record a variety of health parameters. One example is electrocardiography (ECG), which is suitable to detect myocardial infarctions (MIs). Cardiovascular diseases, including MI, are the main cause of death worldwide with more than 17 million deaths every year [1]. MI requires an immediate response as every minute delay reduces the chances of fully recovering. Depending on the severity of the MI, affected persons are unable to call for help themselves or might not even notice the MI when the severity is low.

Smart wearables enable continuous and unobtrusive health monitoring, but the continuous measurement of vital parameters produces much data, which additionally must be classified in real-time. Deep learning (DL) methods like convolutional neural networks (CNN) are suitable to classify this data accurately and timely.

The analysis with CNNs is complex and requires performant hardware. Smartphones and other mobile devices usually do not provide the required performance or only at expense of high battery drainage. Cloud computing could provide a solution but raise privacy

[1] Corresponding Author: Paulo Haas, E-mail: paulo.haas@plri.de

concerns. Multi-access edge computing (MEC) was introduced with the 5G mobile networks and offers high computing capacity close to the clients [2].
Pham et al. [3] provides an overview of MEC's current and future status. However, previous work (including our own [4]) has not yet considered security and scalability, external devices, or trusted cloud computing [4, 5, 6].

2. Material and Methods

We propose an architecture to analyze smart wearables' signals in the edge (see Figure 1). We evaluated our architecture using a mobile phone with a self-developed application, an edge computing device, and CNN-based analysis.

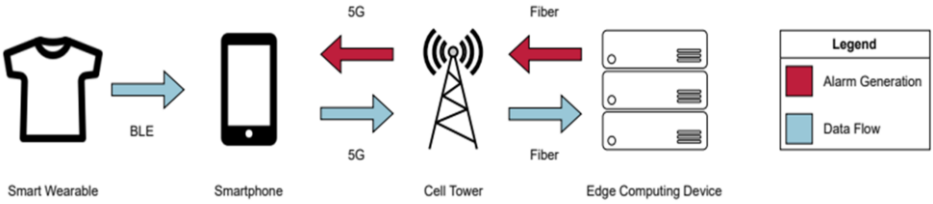


Figure 1. Overview of our proposed architecture.

44 smartphones (Galaxy A52S 5G, Samsung, KOR) generated a simulated ECG signal. All smartphones were equipped with the same business data plan (Business Mobil M, Deutsche Telekom AG, DEU).
A self-developed Android application acquired the data. The app retrieved the current time with the Network Time Protocol (NTP). The smartphone application acted as a Message Queuing Telemetry Transport (MQTT) client and sent the simulated ECG data in batches of eight samples to the server. The Eclipse Paho Android Client library was used for the client implementation.
The edge device was deployed using the far-edge architecture. The data center was chosen based on geographical proximity. *M4.Large* instances served as MQTT brokers (one for secured connection, one for unsecured connection) using Eclipse Mosquitto.
The ECG signals are classified using the 11 layer CNN described in Archarya et al. [7] implemented with Python 3, Tensorflow and Keras and were evaluated with non-specialized hardware without GPU-support. This CNN is provided with single-lead ECG signals and decides binarily whether a MI is detected.

3. Experiment

The experiment aims to analyze 5G for the proposed architecture with respect to latency, data corruption, package loss, inference duration, and energy consumption in consideration of scalability and security-features. The experiment switched between blocks with the Transport Layer Security (TLS)-feature enabled and with TLS disabled. The number of clients decreased logarithmically with three steps per decade: 44, 21, and 9. Values were rounded to the nearest integer. Each block was 15 min long, totaling up to 90 min excluding pauses for reconfiguration and preparation of the next testing block.

4. Results

The study proceeded without any unexpected incidents (e.g.\ application crashes). The system load was low during the experiment. Cell tower connection was monitored by logging their IDs during the experiment to ensure comparability of the data. The study's parameters are evaluated by comparing the client- and server-side log files.
The latencies are computed by aligning the ECG values received on the edge with the values sent by the client and calculating the differences between their timestamps. The latencies are approximately Gaussian distributed (cf. Figure 2). Latency was lowest for nine participants and TLS enabled and highest for 44 participants and TLS disabled.

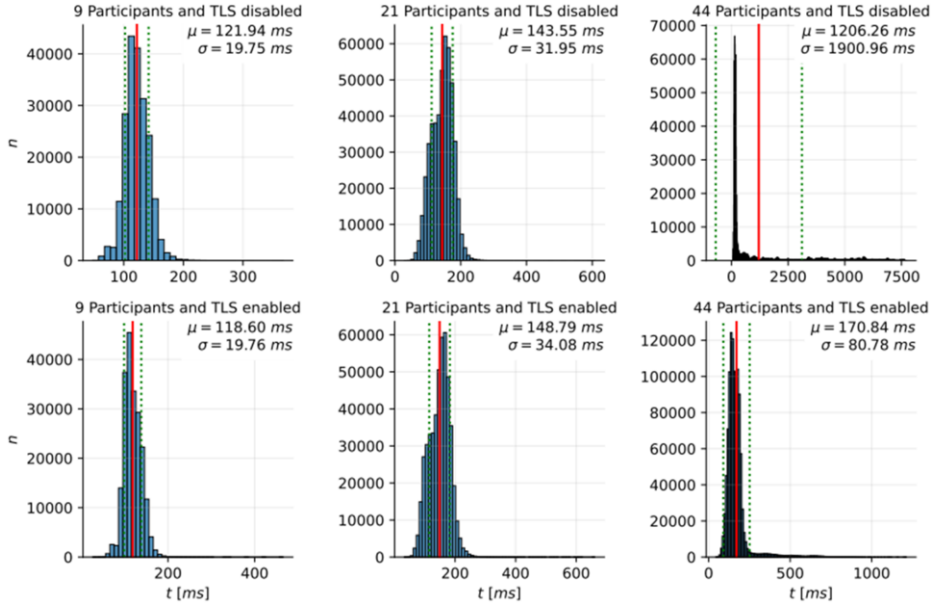


Figure 2. Latencies with their means and standard deviations of transmission from the smartphone to the edge computing device

The data corruption is calculated by aligning the sent and received data. A sliding window can detect missing or unequal values. However, no data corruption was detected in the experiment.
Package loss is calculated by comparing the number of packages sent to the number of packages received. Package loss was lowest for nine participants and TLS disabled and highest for 44 participants and TLS disabled (cf. Table 1).

Table 1. Percentual package loss over the experiment

	9 Participants	21 Participants	44 Participants
TLS enabled	0.001 ± 0.004	0.005 ± 0.013	0.029 ± 0.044
TLS disabled	0.000 ± 0.000	0.002 ± 0.006	0.041 ± 0.079

The inference durations are measured after the experiments. The pre-trained model is fed with the ECG data obtained during the experiment. This data was preprocessed and upsampled according to the algorithm's description [7]. The durations are bimodally distributed (see Figure 3). For all instances the first peak is located within the first few milliseconds and the second peak at approximately 20 ms.

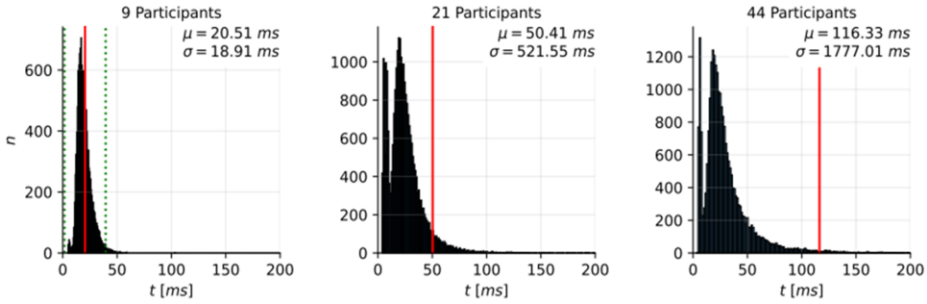


Figure 3. Inference duration with their means and standard deviation.

The energy consumption is logged during the experiments. The battery drainage is 3.221 ± 0.613 % per 15 min when TLS is turned off and 2.978 ± 0.687 % per 15 min when TLS is turned on.

5. Discussion

Our experiments show that end-to-end delays of approximately 500 ms can be realistically achieved even with many patients and encrypted transmissions. This consideration regards that some smart wearables use Bluetooth Low Energy, which achieves delays of up to 50 ms in the worst case [8]. The analysis can be done reliably with low data corruption and low loss, which is below one heart beat per 15 min (0.37 s of loss) and thus clinically irrelevant.

The experiments were conducted with off-the-shelf non-optimized configurations. This essentially means that no other apps ran on the smartphone that could have produced system load or network load influencing the results of our experiment. Also tuning many parameters like the batch size could improve performance without noteworthy drawbacks.

The DL model was trained with a public database, however, this might not be ideal for the detection of MI recorded by mobile ECG sensors. Preprocessing of the data will account for additional delays, which were not considered. The classification of the data then, though evaluated with non-specialized hardware, is quick but GPU-supported classification can further reduce this delay.

The location of the experiment was fixed and the influence of cell tower changes or dead spots was not analyzed concretely. Our proposed architecture will not work in these scenarios and requires fallback solutions like client-side classification, which would cause higher energy consumption and increase inference duration. Third-person devices could have produced concurrent network load negatively influencing our results, which were not compensated. Especially the anomalous distribution of the latency for 44 participants and TLS disabled might be due to cell tower changes. The phones connected to the same tower for only circa 80 % of the time.

Smart wearables providing more sensors could simulate higher data loads and enable other use cases like activity recognition [9], fall detection [10, 11], or detecting deteriorating clinical conditions using the respiration sensors [12].

Finally, 44 patients should not cause high data loads on the mobile network as 5G will support up to 10^6 connections per m^2 [13], but experiments with more volunteers were not possible because of lacking number of devices. The results show that the increasing

number of participants only had little impact on the outcomes. Using more multi-lead ECG sensors would also produce further data. This work evaluated 5G in non-stand-alone mode, which builds upon existing 4G Infrastructure. Future infrastructure will support new modes like 'ultra-reliable low latency communication' or by using network slicing [14], the latencies will be reduced further.

6. Conclusion

Smart wearables record many vital parameters, which can be analyzed by MEC in real-time and with respect to privacy concerns. The proposed MEC platform analyzes signals within 500 ms when evaluated with MI detection by single-lead ECG signal continuously produced by a smart wearable. Future releases of 5G mobile networks could further reduce the time required for classification. Besides, we will repeat the experiment to understand the reasons behind the results more thoroughly.

Appendix

The German Federal Ministry of Transport and Digital Infrastructure (BMVI) funded this work under grant #VB5GFWOTUB.

References

[1] World Health Organization. Health topics - Cardiovascular diseases. [online, accessed 2023-01-15].
[2] Dahlman E, Mildh G, Parkvall S, et al. 5G radio access. Ericsson Rev. 2014 Jun;81(1):42-7.
[3] Pham QV, Fang F, Ha VN, et al. A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art. IEEE Access. 2020 Jun;8:116974-7017.
[4] Spicher N, Klingenberg A, Purrucker V, et al. Edge computing in 5G cellular networks for real-time analysis of electrocardiography recorded with wearable textile sensors. In: Proc 43rd EMBC. IEEE. 2021. p. 1735-9.
[5] Heaney J, Buick J, Hadi MU, et al. Internet of Things-based ECG and vitals healthcare monitoring system. Micromachines. 2022 Dec;13(12):2153.
[6] Liu C, Chen G, Yuan X, et al. Real-time health monitoring system based on wearable devices. In: Proc 16th IWCMC. IEEE. 2020. p. 2002-4.
[7] Acharya UR, Fujita H, Oh SL, et al. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inform Sci. 2017 Nov;415-416:190-8.
[8] Rondón R, Gidlund M, Landernäs K. Evaluating bluetooth low energy suitability for time-critical industrial IoT applications. Int J Wireless Inf Networks. 2017 may;24(3):278-90.
[9] Klingenberg A, Purrucker V, Schuler W, et al. Human activity recognition from textile electrocardiograms. In: Proc 44th EMBC. IEEE. 2022. p. 2434-7.
[10] Stampfler T, Elgendi M, Fletcher RR, et al. Fall detection using accelerometer-based smartphones: Where do we go from here? Front Public Health. 2022 Oct;10:1-13.
[11] Wang Z, Ramamoorthy V, Gal U, et al. Possible life saver: A review on human fall detection technology. Robotics. 2020 Jul;9(3):1-19.
[12] Widiyasari D, Ahmadi N, Adiono T, et al. Contactless respiratory rate monitoring using FMCW radar. In: Proc 6th ISESD. IEEE. 2022. p. 1-6.
[13] Ahad A, Tahir M, Yau KLA. 5G-based smart healthcare network: Architecture, taxonomy, challenges and future research directions. IEEE Access. 2019 Jul;7:100747-62.
[14] Kapassa E, Touloupou M, Mavrogiorgou A, et al. An innovative eHealth system powered by 5G network slicing. In: Proc 6th IOTSMS. IEEE. 2019. p. 7-12.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230327

AI-Based Gut-Brain Axis Digital Twins

Stéphane MEYSTRE[a,1], Ruud van STIPHOUT[a], Annelies GORIS[a] and Santiago GAITAN[a]
[a] *Imec within OnePlanet Research Center, Nijmegen & Wageningen, The Netherlands*
ORCiD ID: Stéphane Meystre https://orcid.org/0000-0002-7632-9625

Abstract. More than 40% of the adult population suffers from functional gastrointestinal disorders, now considered disorders of the "gut-brain axis" (GBA) interactions, a very complex bidirectional neural, endocrine, immune, and humoral communication system modulated by the microbiota. To help discover, understand, and manage GBA disorders, the OnePlanet research center is developing digital twins focused on the GBA, combining novel sensors with artificial intelligence algorithms, providing descriptive, diagnostic, predictive or prescriptive feed-back.

Keywords. Gut-Brain Axis, Digital Twin, Artificial Intelligence, Smart sensors

1. Introduction

The gut-brain axis (GBA) enables gastrointestinal homeostasis, but also has numerous effects on mood and mental health, cognitive functions, inflammation, allergy, neurological diseases, and metabolic health in general. Recent international studies show that more than 40% of the adult population suffers from functional gastrointestinal disorders [1], now considered disorders of "gut-brain" interactions. These interactions are intertwined in a complex bidirectional neural, endocrine, immune, and humoral communication system modulated by bidirectional signaling with the microbiota [2]. Although ongoing research is amassing strong evidence of the importance of these interactions, its intrinsic complexity poses a great challenge for its understanding and the development of precision health applications. A digital twin (or *virtual* twin) is a virtual representation of a real-world physical system or process (i.e., *physical* twin) with (near) real-time connectivity between the physical and the digital twin, and bidirectional flow of data, information and insights [3]. Digital twins of the GBA (DTGBA) could enhance the way we prevent and treat gastrointestinal disorders, allow us to influence our own behavior by providing us with an actual, continuous, and predictive status of the GBA. However, the GBA complexity poses a major challenge to the implementation of DTGBAs. The multiplicity of factors and scales in the biology of the GBA has prevented science to fully explain all involved processes mechanistically; sensors and actuators alone are not enough to implement a DTGBA. What if we could model some of these mechanisms in a data-driven fashion using artificial intelligence (AI) algorithms (e.g., deep and shallow machine learning, hybrid machine learning, Bayesian inference, mixed-effects models)? This is the question we are addressing at the OnePlanet Research Center (OnePlanet), a partnership between imec, Radboud University, Radboud University Medical Center, and Wageningen University and Research in the Netherlands.

[1] Corresponding Author: Stéphane Meystre, E-mail: stephane.meystre@imec.nl

2. Methods

At OnePlanet, innovative sensors are developed to capture data on multiple biomarkers along the gastrointestinal tract (ingestible sensing and sampling sensors, smart toilet seat and urine sensors, wearables, food ingestion sensors, and Apps for human interactions). A digital health platform combines these data with lifestyle, nutrition, sleep, and stress data. These different sources of data are then combined and analyzed with various AI algorithms to derive information and new knowledge applied to early detection and disease prevention. The resulting insights are delivered in tailor-made interfaces, enabling bidirectional interactions between one's GBA and its virtual representation, allowing to keep track of the historical and real-time status of the GBA functions, recognize patterns, predict outcomes, and prescribe personalized measures to improve GBA, mental, and overall health in an interactive fashion.

3. Results, Discussion and Conclusions

In a DTGBA, novel sensors will provide data about the gastrointestinal tract function, microbiota, nervous system activity, and food or liquids consumed, and an App captures information about symptoms, mood, and the menstrual cycle. After the sensing interface combines these data, AI algorithms deliver descriptive analyses provided back to the physical twin (i.e., the person sensed) for monitoring. Further data processing based on AI algorithms offer diagnostic capabilities (e.g., indicating a possible flare up of inflammatory bowel syndrome), predictive capabilities (e.g., how symptoms will evolve if consuming certain food), and prescriptive capabilities (e.g., providing advice to manage symptoms) as actuating feed-back to the physical twin.

The digital twin platform applied to the GBA offers promising precision health abilities based on novel sensing and AI-based data analysis in a model allowing for (near) real-time feedback to help people prevent, detect, monitor, understand, and manage functional gastrointestinal disorders. This effort is limited to a selection of functions and components of the GBA, with a vision to progressively expand its coverage and capabilities. As powerful technology applied to human health, with tremendous potential but also concerns about possible unintended or unanticipated effects, research ethics and privacy protection are central in our work. In addition to performing trials assessed by research ethics committees, we strive for safe, effective, just, unbiased, and patient-centred AI applications, a complex issue we assess and address using principles for trustworthy (i.e., lawful, ethical, and robust) "responsible AI."

References

[1] Sperber AD, Bangdiwala SI, et al. Worldwide Prevalence and Burden of Functional Gastrointestinal Disorders, Results of Rome Foundation Global Study. Gastroenterology. 2021;160:99-114.e3.

[2] Cryan JF, O'Riordan KJ, Cowan CSM, et al. The Microbiota-Gut-Brain Axis. Physiological Reviews. 2019;99:1877–2013.

[3] Jones D, Snider C, Nassehi A, et al. Characterising the Digital Twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology. 2020;29:36–52.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230328

Associations Between Engagement with the BitHabit Digital Lifestyle Intervention and Changes in Type 2 Diabetes Risk Factors

Hilkka LIEDES[a,1], Elina MATTILA[a], Anita HONKA[a], Pilvikki ABSETZ[b], Kirsikka AITTOLA[c], Suvi MANNINEN[c], Niina LINTU[c], Ursula SCHWAB[c,d], Aino-Maija ELORANTA[c,d], Adil UMER[a], Tanja TILLES-TIRKKONEN[c], Reija MÄNNIKKÖ[c,e], Ilona RUOTSALAINEN[a], Mark VAN GILS[b], Jaana LINDSTRÖM[f], Timo LAKKA[c,d,g], Jussi PIHLAJAMÄKI[c] and Anna-Leena VUORINEN[a]

[a] *VTT Technical Research Centre of Finland Ltd., Finland*
[b] *Tampere University, Finland*
[c] *University of Eastern Finland, Finland*
[d] *Kuopio University Hospital, Finland*
[e] *Terveystalo Healthcare, Finland*
[f] *Finnish Institute for Health and Welfare, Finland*
[g] *Kuopio Research Institute of Exercise Medicine, Finland*

Abstract. Type 2 diabetes (T2D) can be prevented or delayed through a healthy lifestyle. Digital behavior change interventions (DBCIs) may offer cost-effective and scalable means to support lifestyle changes. This study investigated associations between user engagement with a habit-formation-based DBCI, the BitHabit app, and changes in T2D risk factors over 12 months in 963 participants at risk of T2D. User engagement was characterized by calculating use metrics from the BitHabit log data. User ratings were used as a subjective measure of engagement. The use metrics and user ratings were the strongest associated with improvements in diet quality. Weak positive associations were observed between the use metrics and changes in waist circumference and body mass index. No associations were found with changes in physical activity, fasting plasma glucose, or plasma glucose two hours after an oral glucose tolerance test. To conclude, increased use of the BitHabit app can have beneficial impacts on T2D risk factors, especially on diet quality.

Keywords. Diabetes, digital health intervention, effective engagement, eHealth

1. Introduction

Type 2 diabetes (T2D) affects 6.3% of the world's population of all ages, and its prevalence is increasing [1] even though it could be prevented or delayed through a healthy lifestyle. Digital behavior change interventions (DBCIs) may provide scalable and cost-effective means to support lifestyle changes [2]. Some minimum amount of engagement with a DBCI is required for the intervention to invoke beneficial health

[1] Corresponding Author: Hilkka Liedes, VTT Technical Research Centre of Finland Ltd., P.O. Box 1100, FIN-90590 Oulu, Finland; E-mail: hilkka.liedes@vtt.fi.

outcomes. However, sufficient amount and type of engagement, i.e., effective engagement, in different settings is not completely understood [3]. Previously, we developed a DBCI, the BitHabit app, containing a large library of healthy small actions (habits), to support adoption and maintenance of healthy habits among people at risk of T2D [4]. The current study aimed to investigate associations between user engagement with the BitHabit app and changes in T2D risk factors.

2. Methods

This study included 963 participants from the StopDia study which evaluated the effectiveness of the BitHabit app in a 12-month randomized controlled trial in a population at risk of T2D (55±10 years [mean ± SD], 78% female) [4]. User engagement was characterized by calculating use metrics from the BitHabit log data: number of use days during the first month and study period and number of days with reported habit performances. User ratings after one to two month's use were used as a subjective measure of engagement. Outcome variables included changes in diet quality (DQ), physical activity (PA), waist circumference (WC), body mass index (BMI), fasting plasma glucose (FPG), and plasma glucose two hours after an oral glucose tolerance test (2h PG). Magnitude of associations between user engagement and changes in T2D risk factors was evaluated with linear regression, adjusted for baseline value of the outcome. The StopDia study was approved by the Research Ethics Committee of the Hospital District of Northern Savo (Statement 467/2016) and was conducted according to the Declaration of Helsinki and the Responsible Conduct of Research by the Finnish Advisory Board on Research Integrity.

3. Results and Discussion

The BitHabit app was used on a median of 5 (inter-quartile range, IQR, 3-10) days during the first month and 40 (IQR 12-71) days during the study period. The participants reported on a median of 375 (IQR 49-1149) performed habits. All user engagement metrics and user ratings were positively associated with DQ. Number of use days during the study period was also positively associated with WC and BMI. No associations were found with changes in PA, FPG, or 2h PG. In conclusion, an increased use of the BitHabit app can have beneficial impacts on T2D risk factors, especially on DQ.

References

[1] Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi A. Epidemiology of Type 2 Diabetes – Global Burden of Disease and Forecasted Trends. J Epidemiol Glob Health 2020;10(1):107-111, doi: 10.2991/jegh.k.191028.001.

[2] Grock S, Ku JH, Kim J, Moin T. A Review of Technology-Assisted Interventions for Diabetes Prevention. Curr Diab Rep 2017;17(11):107, doi: 10.1007/s11892-017-0948-2.

[3] Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement with Digital Behavior Change Interventions. Am J Prev Med 2016; 51(5):833-842.

[4] Lakka TA, Aittola K, Järvelä-Reijonen E, Tilles-Tirkkonen T, Männikkö R, Lintu N, et al. Real-world effectiveness of digital and group-based lifestyle interventions as compared with usual care to reduce type 2 diabetes risk – A stop diabetes pragmatic randomised trial. Lancet Reg Heal - Eur 2023;24:100527, doi: 10.1016/j.lanepe.2022.100527.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230329

Patients' Introduction to Online Video Consultations in Primary Healthcare

Irene MULI[a,1], Marina TALOYAN[b], Lovisa Jäderlund HAGSTEDT[b] and Maria HÄGGLUND[a]
[a] *Participatory eHealth and Health Data Research Group, Dept of Women's and Children's Health, Uppsala University, Sweden*
[b] *Academic Primary Healthcare Centre, Region Stockholm, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Sweden*
[c] *Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Sweden*

Abstract. This study aimed to illustrate ways primary healthcare patients were introduced to video consultations via the public online care application *Alltid öppet* in Region Stockholm, Sweden. The majority of patients were introduced to this by their providers or other healthcare professionals.

Keywords. Video consultations, Introduction, Primary healthcare

1. Introduction

During the Covid-19 pandemic use of online video consultations (VCs) have increased in the world [1]. Users are often younger, women and have high income [2]. Understanding how patients are introduced to VCs might increase our understanding of why certain groups are more likely to use these services.

Alltid öppet is an online care application that can be used for video consultations and to chat with healthcare professionals in public healthcare in Region Stockholm, Sweden [3]. The aim of this study was to explore ways patients were introduced to the possibility to have VCs through the online care application *Alltid öppet*.

2. Methods

This study is part of a case study investigating patients' and healthcare professionals' experiences of VCs in primary healthcare. As part of the case study a cross-sectional survey was administered to all patients who had a VC through *Alltid öppet* with healthcare professionals at 10 primary healthcare centers included in the case study, during March-May 2022. A total of 4263 patients were invited to participate. The survey consisted of 40 questions. Patients were asked how they learned they could book VCs

[1] Corresponding Author, Irene Muli, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, 1st floor, 752 37 Uppsala, Sweden; E-mail: irene.muli@kbh.uu.se.

with their healthcare provider. The question had five pre-defined options and one free text option. The survey was validated twice using cognitive interviews [4] and pilot tested. Descriptive analysis and content analysis of free text answers were conducted. The study was approved by the Swedish Ethical Review Authority (2021-05096).

3. Results

In total, 514 patients responded to the survey (12% of 4263). The majority of the patients (68 %, n = 346/514) stated that they found out that they could book VCs with their primary healthcare providers through their healthcare provider (Figure 1) followed by ''other'' presented in more detail below. Less common ways were through friends (1%, n = 3/514), posters at the healthcare center (1%, n = 7/514), other advertisement (2%, n = 10/514) and through family (4%, n= 20/514).

Most common ''other'' free text responses include other healthcare professionals, figuring it out themselves, and through the application and websites such as the patient portal 1177.se. Other ways included through work in for example healthcare, through a research study, and through media.

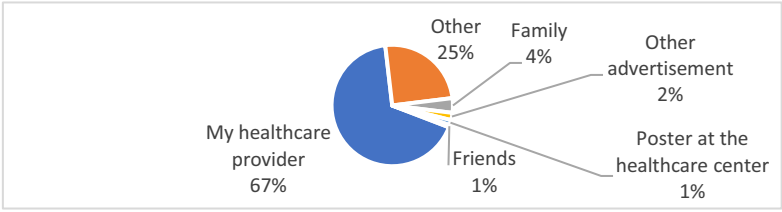


Figure 1. How patients found out about VCs through *Alltid öppet.*

4. Discussion and Conclusions

Social influence is one factor influencing technology acceptance [5], and healthcare professionals' encouragement to use eHealth can increase patients' acceptance and adoption. Further research is needed to determine whether there are differences in how different patient groups are introduced to VCs, and to what extent healthcare professionals' active introduction of the services influences patients' acceptance of VCs.

References

[1] Webster P. Virtual health care in the era of COVID-19. Lancet. 2020 Apr 11;395(10231):1180–1, doi: https://doi.org/10.1016/S0140-6736(20)30818-7

[2] Internetstiftelsen. Svenskarna och internet 2019 [Internet]. Stockholm; Available from: https://svenskarnaochinternet.se/app/uploads/2019/10/svenskarna-och-internet-2019-a4.pdf

[3] Alltid öppet [Internet]. Available from: https://www.vardcentraler.regionstockholm.se/vi-erbjuder/alltid-oppet/

[4] Willis GB, Artino AR. What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. JGME. 2013 Sep;5(3):353–6, doi: https://doi.org/10.4300/JGME-D-13-00154.1

[5] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q. 2003;425–78, doi: https://doi.org/10.2307/30036540

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230330

A Remote Monitoring Platform for the Management of Lower Limb Vascular Diseases

Julio SOUZA[a,b,1], Ana VIEIRA[c], Luís CONCEIÇÃO[c], Rafael MARTINS[c], Daniel RODRIGUES[a], Gustavo CORRENTE[d], William XAVIER[d], Sérgio SAMPAIO[a,b], Goreti MARREIROS[c] and Alberto FREITAS[a,b]
[a] *Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Porto, Portugal*
[b] *CINTESIS@RISE, MEDCIDS, Faculty of Medicine of the University of Porto, Portugal*
[c] *GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Institute of Engineering – Polytechnic of Porto, Porto, Portugal*
[d] *Wiseware Solutions, Portugal*

Abstract. This short paper describes a remote monitoring platform proposed in the Inno4health project. The platform aims to guide patients and clinicians during the treatment of lower limb vascular disorders, namely, to correct abnormal foot pressure and temperature to prevent diabetic foot ulcers and to monitor interface pressure, leg position and elevation for venous ulcers patients.

Keywords. Wearable devices; clinical decision support; vascular diseases; mHealth: remote monitoring; continuous monitoring.

1. Introduction and Methods

Vascular disorders affecting the lower limb, such as diabetic foot ulcers (DFUs) and leg venous ulcers (VUs), can be efficiently addressed by wearable technologies as they require constant follow-up, otherwise leading to considerable loss of quality of life and sometimes the need for hospitalization [1]. The current state-of-the-art includes several systems for continuously monitoring these diseases, but a minority of them enable remote access to patient data. Also, leg VU remains an underexplored field and there are several opportunities to develop innovative wearable devices, namely, to guide routine compression therapy. Automated feedback, recommendations or coaching are also scarce and systems to be used by both, patients and physicians, integrated into the clinical workflow with aid of intelligent services, have not been achieved yet [2]. Thus, one of the aims of the Inno4health project aims is to provide a remote monitoring platform to help professionals and patients detect and actively correct abnormal foot pressure and

[1] Corresponding Author: Julio Souza, E-mail: juliobsouza@med.up.pt.

temperature to prevent DFUs, and monitor pressure under compression apparatus, leg position and elevation in the case of VUs.

A remote monitoring platform targeting the diseases was designed according to requirements defined by literature review [2] and clinicians' feedback obtained through in-situ hospital visits and virtual meetings.

2. Results and Conclusions

The platform's architecture comprises a layer that serves as a gateway with external services and implements secure encrypted connections through the Transport Layer Security protocol. A Data Storage component will maintain pseudonymized data collected from wearable devices, and a medical knowledge base. Data processing components will implement a Rule-Based System to produce and display recommendations for coaching based on the medical knowledge base and run intelligent algorithms for insights generation. A patient's mobile application will be used to collect data from wearable sensors, enable platform interaction and display useful information and recommendations. VU patients will wear a patch to measure leg elevation, motion, time spent in different positions, and pressure under compressive apparatus. The patch's design consists of a pillow with an ultra-slim forcing sensor resistor to measure and log the parameters continuously. DFU patients will wear insoles to measure plantar pressure and temperature in 8 points of the foot to detect areas that can evolve into ulcers.
Physicians will interact with the platform through a web portal, where they can visualize (following authentication) the patient's information and actionable insights in real-time. Regarding DFU, physicians will receive alerts to check the patient whenever persistent localized temperature differences above 1.7˚C between the left and right foot occur or when sustained plantar pressure above 35 mmHg is detected. Regarding VU, clinicians will be able to check if patients are complying with the recommendations on proper leg position and elevation and monitor interface pressure. The prototypes of the insoles, mobile application and web portal for monitoring DFU have been concluded. Future work includes the development and integration of the VU patches, the construction of the medical base for generating personalized recommendations, and pilot studies to receive feedback from patients, which can be further used to improve the technologies.

Acknowledgements: This research work was developed under the project Inno4Health (ITEA-2019-19008-Inno4Health), funded by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (POCI) (ref. POCI-01-0247-FEDER-069523).

References

[1] Kelechi TJ, Johnson JJ, Yates S. Chronic venous disease and venous leg ulcers: An evidence-based update. J Vasc Nurs. 2015;33(2):36-46. doi:10.1016/j.jvn.2015.01.003

[2] Souza, J, et al. Smart Wearable Systems for the Remote Monitoring of Selected Vascular Disorders of the Lower Extremity: A Systematic Review. Int J Environ Res Public Health. 2022 Nov; 19(22), 15231.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230331

Trajectories of Anxiety Symptoms for COVID-19 Patients Using Multimodal Data Collected from Telemedicine

Sumi SUNG[a,b], Su Hwan KIM[b], Youlim KIM[a], Mirinae YOU[a], Hannah LEE[a], Soojin HER[a], Ye Seul BAE[c,1] and Eui Kyu CHIE[a,d,1]
[a]*Office of Hospital Information, Seoul National University Hospital, Seoul, Korea*
[b]*Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea*
[c]*Department of Family Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine*
[d]*Department of Radiation Oncology, Seoul National University College of Medicine, Seoul, Korea*

Abstract. We designed and developed Remote Patient Monitoring (RPM) system specific for coronavirus (COVID-19) patients, and collected multimodal data. Using the collected data, we explored the trajectories of anxiety symptoms for 199 COVID-19 patients quarantined at home. Two classes were identified using latent class linear mixed model. Thirty-six patients showed an exacerbation of anxiety. Presence of initial psychological symptoms, pain on the start day of quarantine, and abdominal discomfort at one month after finishing the quarantine were associated with exacerbation of anxiety.

Keywords. COVID-19, Telemedicine

1. Introduction

Seoul National University Hospital designed a contactless clinical trial [1] and developed remote patient monitoring (RPM) system to expand telemedicine use, and collected multimodal data using online survey, wearable devices, and video call to effectively monitor COVID-19 patients while minimizing contact with the medical staffs. This study aimed to explore the trajectories of anxiety symptoms experienced by COVID-19 patients using the collected data.

2. Methods

A total of 199 adult patients with COVID-19 infection quarantined at home participated in this study from March 2022 to June 2022. The anxiety symptoms were examined using

[1] Co-corresponding Author: Eui Kyu Chie, MD, PhD. Address: Office of Hospital Information, Seoul National University Hospital 101 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea, Email: ekchie93@snu.ac.kr Tel: 82-2-2072-7600, Ye Seul Bae, MD, MS. Address: Department of Family Medicine, Kangbuk Samsung Hospital 29, Saemunan-ro, Jongno-gu, Seoul, Republic of Korea, Email: byeye1313@gmail.com Tel: 82-2-2001-8538

Generalized Anxiety Disorder 7-item scale (GAD-7) [2] at 3 time points: on the start day of quarantine (Time 1), on the final day of quarantine (Time 2), and one month after finishing the quarantine (Time 3). The data was collected from remote patient monitoring system using telehealth. The latent class linear mixed model was used to identify the subgroups of trajectories of anxiety symptoms, and logistic regression was applied with trajectory groups as dependent variables, and sociodemographic features, past medical history, initial psychiatric symptoms, and average number of days with acute COVID-19 symptoms at each of the three time points as independent variables.

3. Results

Two classes of anxiety symptoms' trajectories were identified by sequential GAD-7 score plot of individual patients (Figure 1). Class 1 showed a slow linear declined of anxiety symptoms. Class 2 showed a linear aggravation of anxiety symptoms.

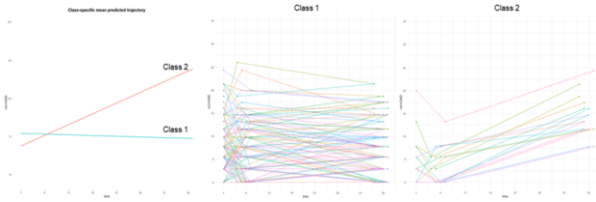


Figure 1. Identified anxiety symptoms' trajectories

Thirty-six patients showed class 2, an exacerbation of anxiety. Presence of initial psychological symptoms (OR, 1.14; 95% CI, 1.00-1.29), pain at Time 1 (OR, 1.12; 95% CI, 1.03-1.22), and abdominal discomfort at Time 3 (OR, 1.23; 95% CI, 1.05-1.45) were associated with exacerbation of anxiety.

4. Conclusion

The results of this study demonstrated the effectiveness of RPM system for COVID-19 patients by expanding the use of telemedicine and possible use to forecast the likelihood of anxiety exacerbation.

Acknowledgement

This work was supported by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-00312.

References

[1] Bae YS, Sung S, Lee J, Lee H, Chie EK. Building and implementing a contactless clinical trial protocol for patients with COVID-19: A Korean perspective. Front Med. 2022 Sep 15;9:975243. doi: 10.3389/fmed.2022.975243.

[2] Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006 May 22;166(10):1092-7. doi: 10.1001/archinte.166.10.1092.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230332

Requirements to mHealth Applications for Animal Owners: A Narrative Review

Laura HAASE[a,1]
[a] *Berlin School of Economics and Law, Germany*
ORCiD ID: Laura Haase https://orcid.org/0000-0003-1632-7334

Abstract. A narrative literature review was conducted to determine software requirements to mHealth applications for animal owners. Results focus on securing a complete user input, providing adequately formulated information to the users and ensuring the understanding of the applications limitations.

Keywords. mHealth, telehealth, animal owners, veterinary medicine

1. Introduction

Like in human medicine, mHealth applications for animal owners may be an approach to relieve veterinarians in their daily work, for example when supporting owners in determining the need of their animal(s) to be seen by a vet [1]. Few research has been conducted that focuses on a technical support solution for symptom recognition and interpretation for animals, whose symptoms need to be recognized and interpreted by their owners as veterinary laypersons in the first place [2]. Existing literature about such applications is scarce and difficult to find, as there are no journals or databases explicitly dedicated to these topics available [3]. This is why a narrative review on requirements to mHealth applications for animal owners is the focus of this paper, being a first step in developing such an application.

2. Methods

A search regarding "veterinary [telemedicine/ telehealth/ mHealth/ informatics]" was carried out in Google Scholar by the author of this paper. The inclusion criteria were ex- or implicitly stated requirements, in German/English language and availability in full text. Publications laying focus on system implementations, ethical/legal/financial aspects, monitoring, specific animal species, concrete veterinary specialties, restricted geographic areas or perceptions/ education of veterinary students were excluded from the review. Duplicates were eliminated and a title/ abstract followed by a full text screening of the most relevant publications were performed. A reference search within all included publications and a review of all telehealth guidelines of veterinary organizations mentioned in one of the full text reviewed publications were carried out. Finally, requirements for implementation were extracted, classified, and summarized.

[1] Corresponding Author: Laura Haase, E-mail: laura.haase@hwr-berlin.de.

3. Results

11 out of the 62 publications reviewed in full text were included in the final analysis. The requirements of particular importance (ordered descending by the number of papers mentioning) are: (i) ensure complete information input by the user [3]; (ii) display a prominent disclaimer about application limitations [4]; (iii) make and display everything in a simple way to facilitate understanding [5]; (iv) display an (emergency) vet contact [6]; (v) use terminology adequate to the user's knowledge [7]; and (vi) display a hint to talk about the application usage with the vet [6].

4. Discussion

So far, there is little specific advice on the implementation of mHealth applications for animal owners. The results reveal that the design of the user interface is crucial. Perhaps this is even more the case than in similar human medicine applications, in order to avoid reporting biases of owners trying to apply their life-learned mental human illness models to the animal. The main limitation of the results, however, is the possible introduction of bias by only one reviewer. Google Scholar was selected as search database in order to draw from an extensive and interdisciplinary pool of publications. Nevertheless, searching in only one database may have left out some publications. Quite broad search terms have been used to gather all papers relevant to the topic, but it may be the case that more specifically named publications have not been covered by the search.

5. Conclusion

Requirements for mHealth applications for animal owners have been elicited from scarce literature; additional original research needs to be carried out. Yet, it seems that veterinary medicine can still benefit from transferrable findings in human medicine specialties like pediatrics. Determining the aspects unique to mHealth applications for animal owners is part of an ongoing research that seeks to develop a successful prototype.

References

[1] Hawk C. Veterinary Telemedicine Perception and Utilization Intention. Knoxville: University of Tennessee; 2018. 47 p.
[2] Weich KL. Chat with your vet! Covid 19 and veterinary tele-medicine. In: Schübel H, Wallimann-Helmer I, editors. Justice and food security in a changing climate; 2021 Jun 24-26; Fribourg, Switzerland. Wageningen (Netherlands): Wageningen Academic Publishers. p. 438.
[3] Smith-Akin K, Bearden C, Pittenger S, Bernstam E. Toward a veterinary informatics research agenda: An analysis of the PubMed-indexed literature. Int J Med Inf. 2007 Apr;76(4):306-312.
[4] Hess L. Telemedicine: The Future of Veterinary Practice. J Avian Med Surg. 2017 Jun;31(2):165-171.
[5] Papageorges M, Hanson J, Girard E, Leveille R, Hebert P, Latouche JS, Menard M. How to make telemedicine work. Clin Tech Small Anim Pract. 2001 May;16(2):111-114.
[6] AVMA Practice Advisory Panel. Final Report on Telemedicine. Whitepaper. 2017 Jan. Available at: https://www.avma.org/sites/default/files/resources/Telemedicine-Report-2016.pdf.
[7] Engelbrecht R, Ingenerf J, Reiner J. Relevance of Terminological Standards and Services in Telemedicine. In: Zielinski K, Duplaga M, Ingram D, editors. Information Technology Solutions for Healthcare. London: Springer; 2006. p. 110-134.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230333

Implementing Remote Patient Monitoring for Patients During Systemic Cancer Therapy

Lara KOHN[a,b,1], Anita MUELLER[c], Veit SCHEBLE[a], Philip STORZ[c], Hanna BORLINGHAUS[b], Selcan Behiye ULAS[c], Nisar MALEK[a, b, d] and Christian THIES[c]
[a] *Clinic of Internal Medicine, University Hospital Tübingen, Germany*
[b] *Center for Personalized Medicine, University Hospital Tübingen, Germany*
[c] *School of Informatics, Reutlingen University, Germany*
[d] *M3 Research Institute, University of Tübingen, Germany*

Abstract. Mobile monitoring of outpatients during cancer therapy becomes possible through technological advancements. This study leveraged a new remote patient monitoring app for in-between systemic therapy sessions. Patients' evaluation showed that the handling is feasible. Clinical implementation must consider an adaptive development cycle for reliable operations.

Keywords. bwHealthApp, Chronic disease, Digital health integration, mHealth

1. Introduction

Remote Patient Monitoring (RPM) has the potential to support personalized cancer therapy monitoring outside of hospitals using wearables to capture outpatients' vital signs continuously [1]. Making RPM available for patient care needs the integration of a software platform into clinical processes. In this work a suitable setting is proposed, and its principal applicability is verified in a pilot study.

2. Methods

The bwHealthApp is used to capture data and patient-reported outcomes.[2] It consists of an Android app for the patients' smartphones and a backend server receiving the recorded data via encrypted communication. Doctors manage patients and data via a web interface. Commercial wearables can be connected to the app via Bluetooth Low Energy. Currently a wrist and an in-ear sensor are supplied for heart rate, temperature, oxygen saturation, and step count. The patient can fill out individually scheduled questionnaires and report personal relevant events. A new patient-centered process for RPM was defined and

[1] Corresponding Author: Lara Kohn, doctoral student at University Hospital Tübingen, E-mail: lara.kohn@med.uni-tuebingen.de.
[2] See details at https://bwhealthapp.reutlingen-university.de/page/home.

implemented at an oncology day clinic at the University Hospital Tübingen (Fig. 1). Its applicability for users and its practicability of data management was verified with 32 oncological outpatients during cancer treatment from July to November 2022.[3]

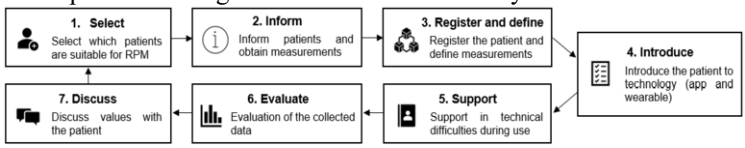


Figure 1. Clinical process for RPM using the bwHealthApp.

After registration and introduction, oncological outpatients should use the app and a wearable for one month. During the first use of the app, participants were observed and interviewed conducting standardized tasks. The actual frequency of use was evaluated, and the system operations for user feedback and software maintenance were examined.

3. Results and Conclusions

During the study, over 20 million values from all subjects were collected. On average, each of the 32 patients generated more than 20.000 data points daily. The frequency of usage was, on average, 14.06 days (SD = 9.35 d) for questionnaires or events and 18.56 days for wearable measures (SD =11.06 d). 14 of 32 patients needed assistance using the bwHealthApp system for the first time. However, 28 of 32 patients stated they had solved at least one task well. Critical comments on the handling of the wearables were expressed by only four participants. Feedback regarding the usability of the app was given. During the period of the study, four releases of the bwHealthApp platform were provided. They consisted of 36 commits for the app and 26 commits for the backend server. The commits included adaptions of user and patient management as well as bug fixes.

The app was, on average, used for half of the observation period. Though individual adherence differed notably. This must be considered in the clinical process, as extended support is needed. The activity of over-average users indicates a willingness to manage technical pitfalls such as battery life despite high symptom burden. During operations, the evolution of the system according to user feedback is possible without limitations based on a suitable DevOps infrastructure. The study showed that ongoing RPM during cancer therapy in times of personalized medicine is feasible but requires a viable integration into the clinical process. For further development and clinical application, feedback on longer use and clinical analysis of the recorded data is needed.

This project is funded by the ministry of social affairs and integration, state Baden-Württemberg, Germany, and realized within the Centers for Personalized Medicine Baden-Württemberg (ZPM). The authors declare that they have no competing interests.

References

[1] Beauchamp UL, Pappot H, Holländer-Mieritz C. The Use of Wearables in Clinical Trials During Cancer Treatment: Systematic Review. JMIR Mhealth Uhealth. 2020;8(11):e22006. doi:10.2196/22006.

[3] Ethical approval received from the University of Tübingen (no. 046/2022BO1), chair Prof. Jaschonek.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230334

Interest in and Experience with the Use of Patient Portals Among Adolescents in Mental Health Care

Martine Stecher NIELSEN [a,1], Aslak STEINSBEKK[b,c] and Torunn Hatlen NØST [a,d]
[a] *Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway*
[b] *Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway*
[c] *Norwegian Centre for E-health Research Tromsø, Norway*
[d] *Norwegian Advisory Unit on Complex Symptom Disorders, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway*
ORCiD ID: Martine Stecher Nielsen https://orcid.org/0000-0002-8088-0559

Abstract. Patient portals hold the potential to support patients and enhance treatment, yet some concerns exist for adults in mental health care and adolescents in general. Due to limited studies on patient portal use in adolescent mental health care, the aim was to examine the interest in and experiences with the use of patient portals among adolescents in mental health care. In a cross-sectional survey, adolescent patients in specialist mental health care across Norway were invited between April and September 2022. The questionnaire included questions on their interests in and experiences with using patient portals. Fifty-three (8,5%), adolescents between 12-18 (mean: 15) responded, of which 64 % were interested in using patient portals. Almost half of the respondents would share access to their patient portal with healthcare providers (48 %) and designated family members (43 %). One-third had used a patient portal, where 28 % had used it to change an appointment, 24 % to see their medications and 22 % to communicate with healthcare providers. The knowledge from this study can be used to inform the set-up of patient portal services for adolescents in mental health care.

Keywords. Patient portals, telehealth, eHealth, electronic health records, adolescent mental health care

1. Introduction

Patient portals may support adolescents in mental health care by e.g., enabling access to health information [1,2]. Research on patient portal use for adolescents outside and adults in mental health care reports both potentials, and concerns e.g., ethical concerns on confidentiality for minors and on access to mental health notes [1,2,3]. Yet limited studies exist on patient portal use for adolescents in mental health care [5]. Thus, knowledge about adolescents' perspectives is needed to inform the setup and use of

[1] Corresponding Author: Martine Stecher Nielsen, Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway, E-mail: martine.s.nielsen@ntnu.no, Tel +4747788730

patient portals for adolescents in mental health care. The aim was to examine the interest and experiences of using patient portals among adolescents in mental health care.

2. Method

A cross-sectional survey was conducted between April and September 2022. A questionnaire was sent to adolescents between 12 and 18 years of age receiving treatment at four specialist child and adolescent mental health services across Norway. They were asked about their interest in and experience with using patient portals.

3. Results

Fifty-three (8,5 %) adolescents between 12-18 years of age (mean 15) responded, of which 64 % were interested in using patient portals. Almost half of the respondents would share access to their patient portal with healthcare providers (48 %) and designated family members (43 %). One-third had used a patient portal, where 28 % had used it to change an appointment, 24 % to see their medications and 22 % for communication.

4. Conclusions

The majority of respondents were interested in using patient portals, and one-third had used a patient portal. The knowledge from this study can be used to inform the set-up and use of patient portal services for adolescents in mental health care.

References

[1] Kipping S et al. A Web-Based Patient Portal for Mental Health Care: Benefits Evaluation. J Med Internet Res. 2016; 18(11).
[2] Leung K et al. Patient and Family Member Readiness, Needs, and Perceptions of a Mental Health Patient portal: A Mixed Method Study. Stud Health Technol Inform. 2019; 266-270.
[3] Sipanoun P et al. A systematic review of the experiences and perceptions of users of an electronic patient record system in a pediatric hospital setting. Int jour of med info. 2022:10469
[4] Schwarz J, Bärkås A, Blease C, Collins L, Hägglund M, Markham S, Hochwarter S. Sharing Clinical Notes and Electronic Health Records With People Affected by Mental Health Conditions: Scoping Review. JMIR Ment Health. 2021 Dec 14;8(12):e34170
[5] Kendrick EJ, Benson C. Patient portals in child and adolescent psychiatry. Child Adolesc Psychiatr Clin. 2017;26(1):43–54.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230335

Automated Classification of Exercise Exertion Levels Based on Real-Time Wearable Physiological Signal Monitoring

Aref SMILEY[a,1], Ihor HAVRYLCHUK[a], Te-Yi TSAI[a], Elena ZAKASHANSKY[a], Aileen GABRIEL[a], Taulant XHAKLI[a], Wanting CUI[a], Irena PARVANOVA[a], Hu CUI[a] and Joseph FINKELSTEIN[a]
[a] *Center for Biomedical and Population Health Informatics, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract. This study aimed to build machine learning (ML) algorithms for the automated classification of cycling exercise exertion levels using data from wearable devices. The best predictive features were selected using the minimum redundancy maximum relevance algorithm (mRMR). Top selected features were then used to build and assess the accuracy of five ML classifiers to predict the level of exertion. The Naïve Bayes showed the best F1 score of 79%. The proposed approach may be used for real-time monitoring of exercise exertion.

Keywords. Aerobic exercise, exertion level, wearable devices, machine learning

1. Introduction and Methods

The goal of this study was to analyze physiological signals collected during an aerobic exercise to automatically identify exertion levels using ML classification techniques. Using our previously developed (iBikE) system [1], ten healthy individuals were asked to perform a pedaling exercise for 16 minutes. For each session, the participants followed the same protocol, which consisted of the first two minutes of pedaling with no resistance, followed by four minutes of pedaling with medium resistance, and ended with the last ten minutes of pedaling with full resistance, when the participants were asked to pedal as fast as they could. The bike resistance was adjusted without interrupting the user's pedaling during each exercise session. Participants were asked to report the "rating of perceived exertion (RPE)" at the end of each consecutive minute of pedaling based on a 1-10 Borg scale of RPE. A total of 16 ratings, one for each minute, were collected for each exercise session. In parallel, the real-time revolutions per minute (RPM) data from the bike, ECG signal from Actiheart 5, pulse rate, and oxygen saturation levels from WristOx2®, Model 3150, were collected. A 2-minute moving window, with one minute shift, was applied to each 16-minute exercise session to divide it into a total of fifteen 2-minute windows. The heart rate variability (HRV) characteristics in time and frequency domains were extracted from the collected ECG signals for each 2-minute window. In addition, collected oxygen saturation levels, pulse rate, and RPMs were averaged for each window. Finally, each 2-minute set of predictive features was assigned to one of 2

[1] Corresponding Author: Aref Smiley, E-mail: Aref.Smiley@gmail.com.

classes based on the corresponding RPE. Class I: high exertion (RPE >= 3.5), Class II: low exertion (RPE < 3.5). As there were no health risks and the participants were all authors of this paper, we did not require institutional review board approval. No protected health information was collected, and the resulting analytical data set was fully de-identified. No compensation was provided to the study participants.

2. Results

We made a dataset consisting of all exercise sessions with all the predictor variables from different users. Then, they were ranked for classification using mRMR algorithm based on the response variables (high exertion vs low exertion). We selected the top six predictors which showed the highest values by mRMR: 1) High-frequency power (ms2), using AR's periodogram estimation methods, 2) The maximum line length inverse (1/Lmax (beats), the divergence), 3) Low-frequency power (%), using Welch's spectrum estimate method, 4) Respiration rate (Hz), 5) Logarithm of very low-frequency power (lo), using Welch's spectrum estimate method, and 6) The RPM. Each predictor consisted of 150 values (fifteen for each user, a total of ten users). Twenty percent of the total values were randomly chosen for testing the trained algorithms. We used MATLAB® Classification Learner app to train models to classify the data. Five major ML techniques were deployed. Table 1 provides a summary of the classifiers' results. The Naïve Bayes technique with Kernel Naïve Bayes classifier had the best performance based on the highest accuracy (80%) and highest F1 score (79%).

Table 1. Results of ML model training and testing using machine learning techniques.

Model	Validation **Accuracy**	Validation **F1 Score**	Validation **AUC**	Test **Accuracy**	Test **F1 Score**	Test **AUC**
Naïve Bayes	78.3%	77%	0.85	80%	79%	0.75
Neural Network	74.2%	74%	0.78	76.7%	72%	0.73
Ensemble	70%	69%	0.81	76.7%	74%	0.83
Kernel	70%	70%	0.69	76.7%	74%	0.77
Tree	79.2%	79%	0.79	73.3%	75%	0.73

3. Discussion and Future work

Automated assessment of exercise exertion levels can potentially improve the safety and efficacy of unsupervised home-based exercise. The developed algorithms must be further tested in large representative groups of patients with different chronic conditions over a prolonged period of time. Impact on exercise adherence, clinical outcomes, and patient safety will have to be established in randomized clinical trials.

References

[1] Smiley A, Tsai TY, Cui W, Parvanova I, Lyu J, Zakashansky E, Xhakli T, Cui H, Finkelstein J. Telemonitoring of home-based biking exercise: Assessment of wireless interfaces. *JMIR Biomed Eng*. 2022;7(2): e41782.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230336

Spectral Fusion of Heartbeat and Accelerometer Data for Estimation of Breathing Rate in Wearable Patches

Mohamed ALHASKIR[a,b,1], Jennifer BAUER[c], Florian LINKE[b], Elisabeth SCHRIEWER[b], Yvonne WEBER[b], Stefan WOLKING[b], Rainer RÖHRIG[a], Markus ROTHERMEL[c], Henner KOCH[b] and Ekaterina KUTAFINA[a]
[a] *Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[b] *Department of Epileptology, Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[c] *Institute for Physiology and Cell Biology, University of Veterinary Medicine Hannover, Hannover, Germany*
ORCiD ID: Mohamed Alhaskir 0000-0002-9983-1554

Abstract. Despite developments in wearable devices for detecting various bio-signals, continuous measurement of breathing rate (BR) remains a challenge. This work presents an early proof of concept that employs a wearable patch to estimate BR. We propose combining techniques for calculating BR from electrocardiogram (ECG) and accelerometer (ACC) signals, while applying decision rules based on signal-to-noise (SNR) to fuse the estimates for improved accuracy.

Keywords. seizures, epilepsy, monitoring, outpatient, electrocardiography, accelerometry, respiratory rate, wearable electronic device

1. Introduction

Chest-worn medical sensors are equipped with both ACC and ECG sensors. The efficacy of both modalities has been previously demonstrated for computing BR [1,2]. Nonetheless, low SNR may lead to imprecise calculation of BR in both cases. Combining estimates from ACC and ECG has shown promising results [3]. Our adaptation of the approach in [3] incorporates an improved BR estimation method from ACC data as in [2], derives spectral information from ACC and ECG signals, and utilizes an SNR-based decision rule for final BR estimation.

2. Methods

The presented data were collected using a wearable device at the RWTH Aachen University Hospital. One-minute recordings were obtained from a patient with epilepsy, during periods of movement and non-movement. BR was determined using respiratory

[1] Corresponding Author: Mohamed Alhaskir, E-mail: malhaskir@ukaachen.de.

sinus arrhythmia (RSA), which corresponds to the modulation of the time intervals between R peaks on ECG during respiration. To derive RSA, R-peaks were detected, and the intervals between them were computed. As in [1], three RSA waveforms were produced by applying three finite impulse response (FIR) bandpass filters of 0.1-0.6 Hz, 0.2-0.6 Hz, and 0.1-0.5 Hz, followed by FFT (Fast Fourier transform). The RSA waveform with a BR (at peak frequency) within 3 bpm of the previous 60-second window was selected. To estimate BR from ACC data, an adaptive line enhancer (ALE) based on least-mean-square (LMS) was initially applied. Following ALE, singular spectrum analysis (SSA) was performed. For each ACC axis, a trajectory matrix was created, followed by singular value decomposition (SVD) to extract two narrow-band signals [2]. Finally, if the mean signal-to-noise ratio (SNRm) of all analyzed signals was > 2dB, majority voting was applied, and the final BR was estimated using the mean BR of all signals except those with a difference in BR > λ (defined by the frequency resolution of FFT) to 40% of the other signals. For SNRm < 2dB, power voting was applied. Here, signals with low magnitude at the frequency of interest were filtered out based on whether the difference between themselves and the signal with the highest magnitude was greater than λ, and the final BR was computed using the mean BR of the remaining signals.

3. Results

The selection of the voting method for estimating BR was based on the SNRm values, which were 3 dB for non-movement segment and -15 dB for movement segment. Thus, majority voting was employed for the former, while power voting was employed for the latter. The resulting BR were 23.74 BPM and 16.5 BPM for the non-movement and movement segments, respectively.

4. Discussion

In this work we present a proof of concept for BR monitoring that builds on state-of-the-art algorithms [1,2,3] but is also adapted to work with data from a specific wearable used for seizure detection in patients with epilepsy. By deriving BR from ACC and ECG data with differences not greater than a conservative threshold of 0.02 Hz, we provided a preliminary clue for the potential functionality of our method. To confirm this, our next step is to test and validate our approach on a larger cohort of patients.

References

[1] Orphanidou C, Fleming S, Shah SA, Tarassenko L. Data fusion for estimating respiratory rate from a single-lead ECG. Biomedical Signal Processing and Control. 2013;8(1):98–105. doi: 10.1016/j.bspc.2012.06.001.

[2] Jarchi D, Rodgers SJ, Tarassenko L, Clifton DA. Accelerometry-based estimation of respiratory rate for post-intensive care patient monitoring. IEEE Sensors Journal. 2018;18(12):4981–9. doi: 10.1109/JSEN.2018.2828599.

[3] Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013. doi: 10.1109/EMBC.2013.6610436.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230337

Voxel-Wise Medical Imaging Transformation and Adaption Based on CycleGAN and Score-Based Diffusion

Feifei LI[a,1], Mirjam SCHÖNECK[b], Oya BEYAN[a] and Liliana Lourenco CALDEIRA[b]
[a] *Institut für Biomedizinische Informatik Köln, Germany*
[b] *Institut für Diagnostiche und Interventionelle Radiologie Uniklinik Köln, Germany*
ORCiD ID: Feifei Li https://orcid.org/0000-0003-4815-8547, Mirjam Schöneck https://orcid.org/0000-0002-1630-4012, Oya Beyan https://orcid.org/0000-0001-7611-3501, Liliana Lourenco Caldeira https://orcid.org/0000-0002-9530-5899

Abstract. Supervised methods, such as those utilized in classification, prediction, and segmentation tasks for medical images, experience a decline in performance when the training and testing datasets violate the i.i.d (independent and identically distributed) assumption. Hence we adopted the CycleGAN(Generative Adversarial Networks) method to cycle training the CT(Computer Tomography) data from different terminals/manufacturers, which aims to eliminate the distribution shift from diverse data terminals. But due to the model collapse problem of the GAN-based model, the images we generated suffer serious radiology artifacts. To eliminate the boundary marks and artifacts, we adopted a score-based generative model to refine the images voxel-wisely. This novel combination of two generative models makes the transformation between diverse data providers to a higher fidelity level without sacrificing any significant features. In future works, we will evaluate the original datasets and generative datasets by experimenting with a broader range of supervised methods.

Keywords. CT, Machine Learning methods, CycleGAN, Score-based Generative Model

1. Introduction

The potential emergence of a distribution shift during machine learning methods execution on different medical image datasets may impede their performances. All these are because supervised machine learning methods must follow its i.i.d assumption. In the study of MRI manufacturer shift and adaptation [1], they used CycleGAN [2] to reverse the distribution shift from different data manufacturers. UT Austin's research team [3] demonstrated the score-based diffusion model can be applied to solving the inverse problem in medical imaging and accelerating and improving the MRI. However, both above methods have limitations; in this study, we successfully cycle-transformed lung cancer CT data from two different terminals and applied the score-based generative model to solve the artifacts problem.

[1] Corresponding Author: Feifei Li, E-mail: feifei.li@uk-koeln.de.

2. Methods

The methods adopted in our work are divided into two processes: the first step is adopting Cycle training to generate coarse scans with patch boundaries, and the second step is utilizing a score-based diffusion model to eliminate the artifacts and enhance the scans to a higher fidelity level.

3. Results

Following the application of cycleGAN, our voxel-wise medical data was generated. The training dataset was 216 images obtained from two CT terminals, namely IMR (Iterative Model Reconstruction) and YA iDose (Intelligent Dose reconstruction with YA kernel). Upon zooming in on the image data, patch boundaries were identified, posing significant diagnostic process challenges. The DDGM [4,5] refining process and corresponding experimental outcomes indicate that it can yield superior synthesis results.

4. Discussion

Based on our experiments, we have observed that both CycleGANs and DDGM methods have shown good performance in certain conditions. CycleGANs were successful in transferring data across different terminals. However, for smaller patches, it suffered from model collapse due to an imbalance in the adversarial training process. The DDGM method is capable of reversing the imaging process, but it is computationally expensive to run the reverse SDE process.

5. Conclusion

Generative models are vital to ensure the stable performance of machine learning methods in medical image analysis, irrespective of the distribution of the input images. In this study, we employed two distinct generative models, namely CycleGAN and score-based DDGM, to address this challenge. Additionally, we plan to evaluate the generated images both qualitatively and quantitively.

References

[1] Yan W, Huang L, Xia L, Gu S, Yan F, Wang Y, et al. MRI Manufacturer Shift and Adaptation: Increasing the Generalizability of Deep Learning Segmentation for MR Images Acquired with Different Scanners. Radiology: Artificial Intelligence. 2020;2(4):e190195. PMID: 33937833

[2] Zhu J, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. CoRR. 2017;abs/1703.10593. Available from: http://arxiv.org/abs/1703.10593.

[3] Jalal A, Arvinte M, Daras G, Price E, Dimakis AG, Tamir JI. Robust Compressed Sensing MRI with Deep Generative Priors. CoRR. 2021;abs/2108.01368. Available from: https://arxiv.org/abs/2108. 01368.

[4] Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:201113456. 2020.

[5] Hyvärinen A, Dayan P. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research. 2005;6(4).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230338

Effects of Ankle-Foot Orthosis on Balance of Foot Drop Patients

Federica AMITRANO[a,b,1] and Armando COCCIA[a,b], Gaetano PAGANO[a], Arcangelo BIANCARDI[a], Giuseppe TOMBOLINI[c] and Giovanni D'ADDIO[a]
[a] *Scientific Clinical Institutes ICS Maugeri SPA SB, Pavia, Italy*
[b] *University of Naples 'Federico II', Naples, Italy*
[c] *Officine Ortopediche Tombolini, Taranto, Italy*

Abstract. Ankle-Foot Orthoses (AFOs) are common non-surgical treatments used to support foot and ankle joint when their normal functioning is compromised. AFOs have relevant impact on gait biomechanics, while scientific literature about effects on static balance is less strong and confusing. This study aims to assess the effectiveness of a plastic semi-rigid AFO in improving static balance on foot drop patients. Results underline that no significant effects on static balance is obtained on the study population when the AFO is used on the impaired foot.

Keywords. Foot Drop, Ankle-Foot Orthosis, AFO, Postural Balance

1. Introduction and Methods

Ankle-Foot Orthosis (AFO) is an external support for foot and ankle joint used when their normal functioning is compromised. Foot drop is one of the most common condition that usually requires the use of an AFO. Orthoses generally improves gait functions and general effects on balance were also registered. However evidence about significant effects of AFOs on balance is less strong, with inconsistent findings due to the different orthoses or tests used [1]. In this study the effects of a semi-rigid AFO on static balance is explored on a group of patients suffering foot drop deficit, whose effects on balance were not deeply explored in literature. The system used to assess balance can provide more specific analytical metrics with respect to previous studies.

The analysis involved twenty-four patients (15 males, age 57±14 yrs, BMI 23.4±4.4 kg/m2), with bilateral (3) or unilateral foot drop syndrome (12 right foot). Static balance assessment was performed using ProKin platform (v. 252 Tecnobody, Dalmine (BG) – Italy). The experimental trial consisted of a static acquisition, repeated with open and closed eyes, in which the subject maintains the standing position for 30s, looking straight forward to a reference point with the feet in extra-rotation. Two-way ANOVA was used to analyse two main effects (open or closed eyes and the influence of the orthosis) and the interaction factor. Statistical analyses were performed using R version 4.0.3.

2. Results, Discussion and Conclusions

Table 1 reports descriptive statistics of the datasets. The ANOVA tests produce the results in Table 2. The level of statistical significance is specified by a different number

[1] Corresponding Author: Federica Amitrano, E-mail: federica.amitrano@icsmaugeri.it.

of asterisks (*). The use of a semi-rigid passive posterior leaf orthosis does not produce improvements in static balance of subjects suffering from foot drop syndrome. The marked differences found between the open- and closed-eyes conditions confirms that this experimental setting can detect changes in static balance functionalities.

Table 1. Postural Parameters for Eyes Open/Closed, in Two Gait Conditions, expressed as mean ±std.

	Without AFO		With AFO	
	Eyes Open	**Eyes Closed**	**Eyes Open**	**Eyes Closed**
Sway Area (cm^2)	4.13±3.58	15.3±21.2	4.69±5.31	17.4±23.8
Sway Path Length (cm)	58.5±29.8	140±114	61.6±34.7	134±96
Standard Deviation AP (mm)	5.97±2.97	10.0±5.7	5.66±2.37	11.0±6.5
Standard Deviation ML (mm)	3.84±2.23	6.28±4.91	4.02±3.03	6.79±5.74
COP Velocity AP (mm/s)	15.3±8.1	39.4±32.9	15.9±9.9	37.7±27.5
COP Velocity ML (mm/s)	8.67±4.79	17.5±14.8	9.25±5.68	16.2±13.4

Table 2. ANOVA Statistical Test Results.

	AFO Condition	**Eyes**	**AFO Condition*Eyes**
Sway Area	0.632	5.37e-5****	0.783
Sway Path Length	0.887	9.32e-10****	0.670
Standard Deviation AP	0.632	2.21e-9****	0.354
Standard Deviation ML	0.579	7.62e-5****	0.794
COP Velocity AP	0.859	2.37e-10****	0.714
COP Velocity ML	0.803	2.31e-6****	0.534

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Several studies have reported enhanced balance confidence in daily activities when using a lower limb orthosis [2], and in quantitative assessment of dynamic balance [3]. However the absence of improvements in static balance is reported in several studies in different patients [4,5]. Probably, the most general conclusion is that AFO principally improves gait and dynamic balance rather than static functions, and that different designs of AFO have different impacts on postural responses, thus being more appropriate for a group of specific patients. Our study focused on a group of subjects suffering from foot drop, and reported no effects on static balance when a passive AFO is used. However, some limitations should be mentioned: the study population is limited and the aetiology of foot drop is varied. Moreover, patients used this orthosis for the first time in this trial and had just few minutes of walking to be confident with it. The structure and the working mechanism of the passive orthosis are very simple, reducing the training time needed to achieve an appropriate use. This study aims to explore the immediate effects on balance of using a passive AFO and prompts future works to investigate the impact of long-term use before progressing to general considerations about the effectiveness.

References

[1] Ashkar M, Razeghinezhad R, Moghadasi Chevinlee H, Tavakoli B, Bagherzadeh Cham M. The effects of orthotics device on the balance control in multiple sclerosis patients: A systematic review and meta-analysis. Mult Scler Relat Disord. 2022 Oct;66:104005–104005.

[2] Pourhoseingholi E, Saeedi H. Role of the newly designed Ankle Foot Orthosis on balance related parametersin drop foot post stroke patients. J Bodyw Mov Ther. 2021 Apr;26:501–4.

[3] Jackson K, et al. The Effects of Carbon and Plastic Ankle-Foot Orthoses on Gait, Balance, and Mobility in Persons with Multiple Sclerosis: A Pilot Study. JPO J Prosthet Orthot. 2021 Jul;33(3):227–33.

[4] Parian SS, et al. The Comparison of the Effects of Flexible vs Rigid Ankle-foot orthoses on Balance and Walking Performance in Individuals With Multiple Sclerosis: A Crossover Study. Iran Rehabil J. 2021;19(2):199–206.

[5] Choo YJ, Chang MC. Effectiveness of an ankle–foot orthosis on walking in patients with stroke: a systematic review and meta-analysis. Sci Rep. 2021 Aug;11(1):15879–15879.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230339

On-Site Visualization of Ballistocardiography Data

Urs-Vito ALBRECHT[a,1] and Ulf KULAU[b]
[a] *Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Bielefeld, Germany*
[b] *Smart Sensors Group, Hamburg University of Technology (TUHH), Hamburg, Germany*
ORCiD ID Urs-Vito Albrecht https://orcid.org/0000-0002-8989-6696,
Ulf Kulau https://orcid.org/0000-0002-1662-880X

Abstract. We describe the background, features and functions of a custom application for the acquisition, live presentation, and convenient recording of ballistocardiography data acquired by external accelerometric sensors.

Keywords. software, ballistocardiography, data recording

1. Introduction

During increments 66 and 67, we conducted Ballistocardiography (BCG) experiments aboard the International Space Station (ISS) [1]. This article focuses on software that was developed to allow on-site live monitoring during pre- and post-flight measurements in preparation for the experiments.

2. Method

The hardware used for the BCG measurements has been described in [2,3]. The software is designed to perform BCG measurements quickly and with high data quality. After a requirements analysis and pen-and-paper development, the application prototype was written in C++, using the QT 5.12 framework in Q1.2021. After several iterations of testing and modification, version 1.0 was ready at the end of Q2 2021.

3. Results

The interface functions as a virtual COM port via USB, with the data being sent as a binary stream in a fixed format. The data is received in the software from the interface and written directly to a file without conversion. Thus, in addition to the later decoded

[1] Corresponding Author: Urs-Vito Albrecht, Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany; E-mail: urs-vito.albrecht@uni-bielefeld.de.

data, the original binary data is also recorded, which can be useful for debugging if necessary. The binary data stream is then decoded, splitting the data into frames and thus making all BCG data available as composite objects. This decoded data and corresponding timestamp can also be stored in a CSV file if required. However, most importantly, the data can be visualized in real-time. All BCG data can be viewed graphically in freely selectable configurations. The graphs also offer features such as pause or zoom. Since BCG data is often located near the noise floor, signal processing with filters is usually necessary. These features are directly integrated into the live view. Moving mean and bandpass filters are implemented, and their depth or cutoff frequencies are also freely configurable in real-time. This allows the data to be viewed directly during the measurement, and the display can be optimized by setting filters individually. If everything is configured and the setup for the measurement process is correct, data recording can be started via the record button. In addition, markers can be set to tag events directly in the recorded data during the measurement. Of course, all measured values are timestamped (Unix time stamp format), and the file is once again saved as CSV. In addition to the live view, all recorded data acquired since the software was started can also be displayed visually.

4. Discussion

The Ballistocardiography (BCG) method utilizes high-precision accelerometers to measure the heart's motion on the body surface and to non-invasively obtain physiological information of the cardiac function [4]. For sufficient on-site analysis, customized software was developed. The software can be extended by further features in the future. The replay of collected data is planned, as well as the automatic annotation of signal features for easy collection of training data for neural networks.

5. Conclusions

In combination with external sensors, the software was successfully used to obtain BCG measurements for two ESA astronauts.

References

[1] Drobczyk M, Lubken A, Strowik C, Albrecht UV, Rust J, Beringer J, et al. A Wireless Communication Network With a Ballistocardiography Experiment on the ISS: Scenario, Components and Preflight Demonstration. IEEE Journal of Radio Frequency Identification. 2022;6:258-68.

[2] Kulau U, Rust J, Szafranski D, Drobczyk M, Albrecht UV. A Differential BCG Sensor System for Long Term Health Monitoring Experiment on the ISS. In: 2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS); 2022. p. 85-92.

[3] Kulau U, Rust J, Albrecht UV. Demo: BCG Measurement by differential Sensing in Real-Time. In: 2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS); 2022. p. 75-8.

[4] Albrecht UV, Drobczyk M, Strowik C, Lubken A, Beringer J, Rust J, et al. Beat to BEAT - non-invasive investigation of cardiac function on the International Space Station. Stud Health Technol Inform. 2022 Jun;295:95-9.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230340

Tracking of Nutritional Intake Using Artificial Intelligence

Marko PETKOVIĆ[a,b,1], Joyce MAAS[c], Milan PETKOVIĆ[a]
[a] *Eindhoven University of Technology, The Netherlands*
[b] *Data Science Department, 5M ICT, Serbia*
[c] *Center for Eating Disorders, GGZ Oost Brabant, The Netherlands*
ORCiD ID: Marko Petković https://orcid.org/0009-0000-4918-6027

Abstract. In this short communication paper, we present the results we achieved for automated calorie intake measurement for patients with obesity or eating disorders. We demonstrate feasibility of applying deep learning based image analysis to a single picture of a food dish to recognize food types and make a volume estimation.

Keywords. Nutrition Measurement, Deep Learning, Eating Disorders

1. Introduction

An increasing number of people are affected by obesity and eating disorders. According to the World Health Organization around two billion adults are overweight, more than 650 million are obese, and around 9% of the population is affected by eating disorders. Most of these diseases can severely affect a person and lead to cardiovascular diseases or even death, while they are expensive to treat. To better manage people with these conditions, it is important to have a solution for automatic but accurate nutrition intake tracking. In this paper, we propose a novel approach, which can accurately recognize and estimate the calorie contents of 101 different food types. For this we make use of various Deep Learning techniques which are integrated into our FitSprite Nutrition smartphone app to protect privacy by running inference locally. The solution also provides a web portal used by clinicians or dietitians where they can find an overview of their patients' eating habits. This allows them to manage consumers/patients more efficiently.

2. Methods

For calculating the amount of calories from a food picture, we propose a framework consisting of three Convolutional Neural Networks (CNNs). The three different models we use are: (i) a CNN for food prediction; (ii) a CNN with U-Net architecture for food segmentation, and (iii) a CNN with U-Net architecture for depth map prediction.

To train the model for food prediction, we make use of an EfficientNet model pretrained on ImageNet. We trained the model on the Food-101 dataset [1], which consists of 101 different food classes, with 1000 pictures per food item.

[1] Corresponding Author: Marko Petković, E-mail: m.petkovic1@tue.nl

For the task of food segmentation, we made use of the UEC-FoodPixComplete dataset [2], which contains 10000 food pictures with their pixel-wise segmentation masks for each individual food item in the picture (a different color mask for fries, steak etc.). We trained a U-Net, using the food prediction model as encoder. To record the performance of the model, we used the Intersection over Union metric.

For depth estimation, we trained a U-Net using self-supervised monocular depth prediction. This model was trained on the EPIC-KITCHENS dataset [3], which contains 100 hours of recordings of food preparation.

3. Results

For the food prediction task, we achieved a top-1 accuracy of 0.81 and a top-5 accuracy of 0.94. The image segmentation network achieved an Intersection over Union of 0.91, while combined with the depth estimation network it achieved a mean absolute percentage error of 11.7% in predicting food volume.

4. Discussion

Overall, we see that it is possible to make fairly accurate calorie intake estimations using our proposed models. We also found that by combining the food prediction and segmentation tasks, we were able to achieve a better performance for food segmentation. The performance of our food prediction network is better than current methods which are able to run locally on smartphones [4]. For the image segmentation task, we obtained solid performance by achieving a higher Intersection over Union than the current best food segmentation network [5].

5. Conclusion

In this paper, we have shown that automated tracking of nutrition is feasible. We finalized two pilots with users of a fitness club and patients with binge eating disorder and their clinicians in the Center for Eating Disorders in Helmond in the Netherlands. The pilots demonstrated the benefits for both, users/patients for the automation in the creation of patient dairies, and coaches/clinicians for more efficient patient management.

References

[1] Bossard L, Guillaumin M, Van Gool L. Food-101–mining discriminative components with random forests. In Computer Vision–ECCV 2014 Proceedings, Part VI (pp. 446-461). Springer.
[2] Okamoto K, Yanai K. UEC-FoodPIX Complete: A large-scale food image segmentation dataset. In Pattern Recognition ICPR. Proceedings, Part V 2021 (pp. 647-659). Springer International Publishing.
[3] Damen D, Doughty et al. Rescaling egocentric vision: Collection, pipeline and challenges for epic-kitchens-100. International Journal of Computer Vision. 2022 Jan 1:1-23.
[4] Lo FP, Sun Y, Qiu J, Lo B. Image-based food classification and volume estimation for dietary assessment: A review. IEEE journal of biomedical and health informatics. 2020 Apr 30;24(7):1926-39.
[5] Ando Y, Ege T, Cho J, Yanai K. Depthcaloriecam: A mobile application for volume-based foodcalorie estimation using depth cameras. In MADiMa (pp. 76-81) https://doi.org/10.1145/3347448.3357172.

Section 10

Bioinformatics

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230342

Improving Patient Similarity Using Different Modalities of Phenotypes Extracted from Clinical Narratives

Xiaoyi CHEN[a,b,c,1], Carole FAVIEZ[b,c], Marc VINCENT[a], Sophie, SAUNIER[d], Nicolas GARCELON[a,b,c] and Anita BURGUN[b,c,e,f]
[a] *Data Science Platform, Imagine Institute, Université de Paris Cité, Inserm UMR 1163, Paris, France*
[b] *Inserm, Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris Cité, Paris, France*
[c] *HeKA, Inria Paris, Paris, France*
[d] *Laboratory of Renal Hereditary Diseases, Imagine Institute, Université de Paris Cité, Inserm UMR 1163, Paris, France*
[e] *Hôpital Necker-Enfants Malades, Département d'informatique médicale, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France*
[f] *PaRis Artificial Intelligence Research InstitutE (PRAIRIE), France*
ORCiD ID: Xiaoyi Chen https://orcid.org/0000-0002-7378-5158

Abstract. In the context of medical concept extraction, it is critical to determine if clinical signs or symptoms mentioned in the text were present or absent, experienced by the patient or their relatives. Previous studies have focused on the NLP aspect but not on how to leverage this supplemental information for clinical applications. In this paper, we aim to use the patient similarity networks framework to aggregate different phenotyping modalities. NLP techniques were applied to extract phenotypes and predict their modalities from 5470 narrative reports of 148 patients with ciliopathies (a group of rare diseases). Patient similarities were computed using each modality separately for aggregation and clustering. We found that aggregating negated phenotypes improved patient similarity, but further aggregating relatives' phenotypes worsened the result. We suggest that different modalities of phenotypes can contribute to patient similarity, but they should be aggregated carefully and with appropriate similarity metrics and aggregation models.

Keywords. patient similarity, deep phenotyping, negated phenotype, experiencer

1. Introduction

Medical concept extraction from clinical narratives is an important subdomain of biomedical natural language processing (NLP), enabling applications ranging from clinical decision support to care quality improvement [1]. In addition to concept detection, normalization and disambiguation, extraction of context information, such as negation and experiencer (referred to as "modalities"), is critical for determining whether mentioned clinical signs were present or absent, experienced by the patient or by their

[1] Corresponding Author: Xiaoyi Chen, E-mail: xiaoyi.chen@institutimagine.org

relatives, which can have a high impact in clinical applications. Although recent efforts have improved modality prediction accuracy [2], little attention has been given to using this supplemental information in subsequent work. Previously, we used this metadata to keep only the patient's own present phenotypes when computing patient similarity for rare disease diagnosis. Recently, Slater et al. evaluated this choice in classifying primary diagnosis over MIMIC-III patient visits and confirmed its effectiveness [3].

Similar clinical characteristics are believed to be indicative of similar clinical outcomes [4]. This study hypothesizes that negated phenotypes and family histories can also contribute to patient similarity, as the negation may reflect the assumptions of clinicians, and family history may also be useful in predicting outcomes for health conditions related to genetic inheritance. Therefore, rather than simply removing negated phenotypes and relatives' phenotypes, we aim to aggregate different phenotyping modalities into similarity models. Patient similarity network (PSN) framework has been often considered for heterogeneous multi-omics data aggregation (such as mRNA expression, DNA methylation, etc.) [4], where all types of data are converted to a single type of input (similarity networks), integration is straightforward [5].

In this study, we explored the feasibility of using PSNs framework to improve patient similarity using different phenotyping modalities, namely negated phenotypes and phenotypes experienced by patient's family members. The study was conducted as part of the C'IL-LICO program, aiming to develop transformative diagnostic, prognostic and therapeutic approaches for patients suffering from ciliopathies, a group of rare diseases caused by ciliary dysfunction. The proposed method was evaluated in the context of stratifying ciliopathy patients into subgroups using deep phenotyping in their unstructured electronic health records (EHRs).

2. Materials and methods

2.1. Patient selection and ciliopathies subtypes

The joint data warehouse of Necker Children's Hospital and Imagine Institute, called Dr Warehouse, holds over 9 million documents of 800,000 patients, and structured data (gene, diagnosis, manually curated phenotypes) for more than 1200 patients with ciliopathies. This study focused on the 148 diagnosed patients at Necker Children's Hospital with sufficient documents, involving 47 ciliary genes and 26 Orphanet encoded diagnoses. The gene-diagnosis combination created 64 classes, with 54 classes having less than 3 patients. We thus grouped the genes based on their function and localization within the cilium based on [9]. Overlapping diagnoses were also grouped. The final class assignment was validated by a ciliopathy expert (SS), which resulted in five classes.

2.2. Clinical concept extraction and modality prediction

For phenotype extraction, a hybrid strategy combining a dictionary-based approach and a deep-learning approach using bidirectional Gated Recurrent Units and Conditional Random Fields (biGRU-CRF) model was adopted, representing extracted mentions as concepts in the Human Phenotype Ontology (HPO). For modality prediction, a deep learning pipeline was developed using fastText and contextual Bidirectional Encoder Representations from Transformers (BERT)-type embeddings, combined with GRU or

Long Short Term Memory (LSTM) recurrent neural networks, which were shown outperforming the rule-based approaches in negation and subject prediction tasks [6].

2.3. Patient similarity networks using different modalities of phenotyping

For each phenotype modality, i.e., patient's positive (pt_pos) and negative (pt_neg), family's positive (fm_pos) and negative (fm_neg), patient similarities were computed using the average best match method as described in [7]. More precisely, for two patients represented by two sets of concepts P_1 and P_2, the similarity from patient P_1 to P_2 is the weighted average of the best-match concept similarities over all concepts in P_1:

$$sim_{set}(P_1 \to P_2) = \frac{1}{|\boldsymbol{a_1} P_1|} \sum_{p_{1i} \in P_1} a_{1i} \max_{p_{2j} \in P_2} sim_{concept}(p_{1i}, p_{2j}), \qquad (1)$$

where $\boldsymbol{a_1}$ is the weight vector indicating the relevance of each phenotype to the patient. The symmetric similarity is $sim_{set}(P_1, P_2) = \frac{1}{2}\left(sim_{set}(P_1 \to P_2) + sim_{set}(P_2 \to P_1)\right)$. Regarding the similarity between concepts, we considered the Lin's semantic similarity [8], which is based on the information content (IC) of the two phenotypes and the IC of their lowest common subsumer (LCS) in the HPO hierarchy:

$$sim_{concept}(p_i, p_j) = \frac{2 \times \mathrm{IC}\left(\mathrm{LCS}(p_i, p_j)\right)}{\mathrm{IC}(p_i) + \mathrm{IC}(p_j)}. \qquad (2)$$

Four PSNs were built using the computed patient similarity matrix, with patients as nodes and similarities as weighted edges. The $\mathrm{PSN}_{\mathrm{pt\_pos}}$ was considered as a baseline. Then the other PSNs were aggregated successively by considering a weighted average combination of the current network and previously aggregated networks: $\mathrm{PSN}_{\mathrm{agg2}} = \alpha \mathrm{PSN}_{\mathrm{pt\_pos}} + (1-\alpha)\mathrm{PSN}_{\mathrm{pt\_neg}}$, $\mathrm{PSN}_{\mathrm{agg3}} = \beta \mathrm{PSN}_{\mathrm{agg2}} + (1-\beta)\mathrm{PSN}_{\mathrm{fm\_pos}}$, and so on.

2.4. Clustering and evaluation for patient stratification

Hierarchical agglomerative clustering with complete-linkage was applied on the baseline PSN and on each aggregated PSN. The number of clusters was fixed to 5, which equals to the number of grouped gene-diagnosis classes. Rand index (RI) and adjusted RI (ARI) were used to measure the concordance with the ground truth class assignment. More precisely, $\mathrm{RI} = (\mathrm{a} + \mathrm{b})/C_2^N$, where a is the number of pairs of patients that are in the same gene-diagnosis class and in the same cluster, b is the number of pairs of patients that are in different classes and in different clusters, and C_2^N is the number of all possible pairs of patients. The ARI is the RI discounted by the expected RI of random labelings:

$$\mathrm{ARI} = \frac{\mathrm{RI} - \mathbb{E}(\mathrm{RI})}{\max(\mathrm{RI}) - \mathbb{E}(\mathrm{RI})}. \qquad (3)$$

Therefore, the range be definition of RI and ARI are [0,1] and [-1,1], respectively.

3. Results

3.1. Different phenotype modalities and PSNs

After applying the NLP techniques, we identified 2157 distinct pt_pos phenotypes (extracted from 5470 narrative reports from 148 ciliopathy patients), 879 distinct pt_neg phenotypes (3366 reports of 131 patients), 275 distinct fm_pos phenotypes (575 reports

of 66 patients), and 49 distinct fm_neg phenotypes (124 reports of 21 patients). Due to the small number of patients with fm_neg phenotypes, patient similarities were computed only for the first three modalities, and the aggregation was limited to these three PSNs. Figure 1 showed the PSNs with the Fruchterman-Reingold layout. To facilitate the visualization, only the top 5% of the weighted edges for each node are displayed. Nodes were colored by the ground truth class assignment.

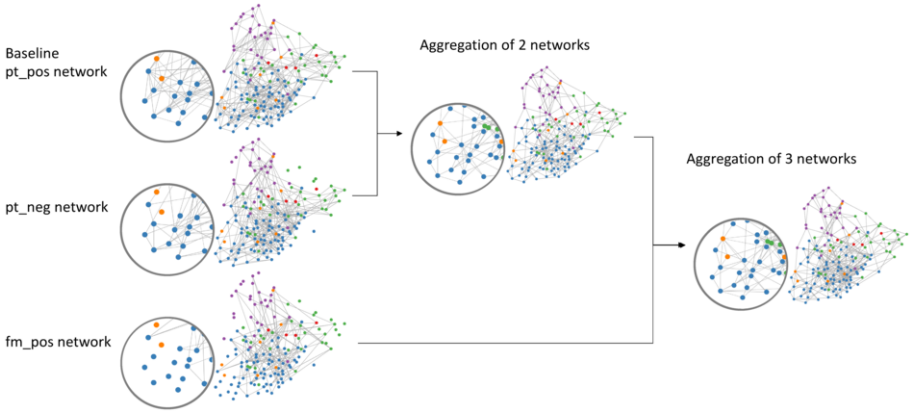


Figure 1. Individual and aggregated networks with a zoom on the bottom left.

3.2. Clustering performance

The RI and ARI were computed for clustering using PSN_{pt_pos} and each aggregated PSN, the results obtained with the best aggregation weights that were determined via grid search for each aggregation are shown in Table 1. We can observe that aggregating patient's negated phenotypes improved the performance, but further aggregating family's phenotypes worsened the result. In order to better understand the performance improvement using PSN_{agg2}, we applied the same clustering method with the same similarity model to manually curated phenotypes in the structured research data. In contrast to the automated phenotype extraction from narrative reports, manually curated phenotypes are comprehensive, precise and relevant, thus can be considered as the best phenotypic representation of patient. The RI and ARI using the manually curated phenotypes was 0.784 and 0.514, respectively, which are not close to 1, due to the phenotypic and genetic heterogeneity and overlap of ciliopathies. Given these values as references, the relative increase in RI and ARI towards the references is 43% and 66%, respectively, showing an important improvement. An expert (SS) reviewed the discrepancies of the clustering results obtained from PSN_{pt_pos} and PSN_{agg2}, and confirmed the improvement by aggregating patient similarity on negated phenotypes. For example, a better stratification of isolate and syndromic Leber congenital amaurosis (LCA) caused by different genes was achieved using PSN_{agg2}.

Table 1. Clustering performance for the baseline PSN (pt_pos), aggregated PSNs, and the PSN using manual curated phenotypes

	Baseline pt_pos	**Agg2 +pt_neg**	**Agg3 +pt_neg+fm_pos**	**Reference manually curated phenotypes**
Rand index	0.735	**0.756**	0.721	0.784
adj. Rand index	0.393	**0.473**	0.349	0.514

4. Discussion

Gliozzo et al. [4] distinguished three aggregation approaches in the recent review: input data-fusion, PSN-fusion, and output-fusion. Input data-fusion is unsuitable for negations and family histories, since the same phenotype can be present and absent for the same patient at different times, and for the patient and also for the relatives. Output-fusion is also not possible, since data in some modalities can be too sparse for a reliable prediction. Therefore, an intermediate integration was considered to compute patient similarities using each modality independently, then aggregating them successively. While the lack of ground truth makes evaluating clustering performance challenging, expert grouped gene-diagnosis classes were established, and RI and ARI using manually curated phenotypes were provided as a reference. Our results showed significant improvement of ARI when aggregating negated phenotypes, but worse performance with family's phenotypes. We think that is because most of the ciliopathies in our dataset are recessive disorders, and the subject prediction can only distinguish whether the experiencer is the patient or not, without further precision (parents, siblings, or other family members).

Our study has some limitations, which will be addressed in future work. As the first attempt of aggregating different phenotyping modalities into patient similarity, we used the same similarity metric for negated phenotypes as for positive phenotypes, and a simple aggregation model, i.e., a weighted sum of individual similarities obtained on each modality. The evaluation was conducted on a small dataset. The next step will be to explore other similarity metrics that may suit better negation and family history, investigate more sophisticated aggregation models, and perform a broader evaluation involving also dominant disorders to assess the impact of family history.

This work was supported by State funding from The French National Research Agency (ANR) under "Investissements d'Avenir" programs (ANR-10-IAHU-01) and C'IL-LICO project (ANR-17-RHUS-0002).

References

[1] Fu S, Chen D, He H, Liu S, Moon S, Peterson KJ, Shen F, Wang L, Wang Y, Wen A, Zhao Y, Sohn S, Liu H. Clinical concept extraction: A methodology review. J Biomed Inform 2020 Sep;109:103526. PMID:32768446

[2] Slater LT, Bradlow W, Motti DF, Hoehndorf R, Ball S, Gkoutos GV. A fast, accurate, and generalisable heuristic-based negation detection algorithm for clinical text. Comput Biol Med 2021 Mar;130:104216. PMID:33484944

[3] Slater LT, Karwath A, Hoehndorf R, Gkoutos GV. Effects of Negation and Uncertainty Stratification on Text-Derived Patient Profile Similarity. Front Digit Health 2021 Dec 6;3:781227.

[4] Gliozzo J, Mesiti M, Notaro M, Petrini A, Patak A, Puertas-Gallardo A, Paccanaro A, Valentini G, Casiraghi E. Heterogeneous data integration methods for patient similarity networks. Brief Bioinform 2022 Jul 18;23(4):bbac207. PMID:35679533

[5] Pai S, Hui S, Isserlin R, Shah MA, Kaka H, Bader GD. netDx: interpretable patient classification using integrated patient similarity networks. Mol Syst Biol 2019 Mar 14;15(3):e8497. PMID:30872331

[6] Vincent M, Douillet M, Lerner I, Neuraz A, Burgun A, Garcelon N. Using Deep Learning to Improve Phenotyping from Clinical Reports. Stud Health Technol Inform 2022 Jun 6;290:282–286.

[7] Chen X, Faviez C, Vincent M, Garcelon N, Saunier S, Burgun A. Identification of Similar Patients Through Medical Concept Embedding from Electronic Health Records: A Feasibility Study for Rare Disease Diagnosis. Stud Health Technol Inform 2021 May 27;281:600–604. PMID:34042646

[8] Lin D. An Information-Theoretic Definition of Similarity. Proc Fifteenth Int Conf Mach Learn San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1998. p. 296–304.

[9] Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol 2017 Sep;18(9):533–547. PMID:28698599

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230343

OmicSDK-Transcriptomics: A Web Platform for Transcriptomics Data Analysis

Aurora SUCRE[a, b, 1], Maria MARTINEZ [b] and Alba GARIN-MUGA [a, b]
[a] *Biodonostia Health Research Institute, Basque Research and Technology Alliance (BRTA), 20014 Donostia-San Sebastián, Spain*
[b] *Fundación Vicomtech, Basque Research and Technology Alliance (BRTA), 20009 Donostia-San Sebastián, Spain*
ORCiD ID: Aurora SUCRE https://orcid.org/0000-0002-4078-9275
ORCiD ID: Alba GARIN-MUGA https://orcid.org/0000-0002-7160-1191

Abstract. Omics sciences, especially transcriptomics, have grown exponentially since the first human genome was sequenced in 2003. Different tools have been developed in the past years for the analysis of this kind of data, but many of them require specific programming knowledge to be used. In this paper, we present omicSDK-transcriptomics, the transcriptomics module of OmicSDK, a comprehensive tool for omics data analysis that combines pre-processing, annotation and visualization tools to be used with omics data. OmicSDK comprises a command-line tool and a user-friendly web solution, so researchers having different backgrounds can take advantage of all its functionalities.

Keywords. Transcriptomics, bioinformatics, data analysis, visual analytics

1. Introduction

Since the sequencing of the first human genome in 2003, omics sciences have grown exponentially. One of the most developed sciences in the field is transcriptomics, which focuses on the study of the transcriptome, the sum of all the RNA transcripts of an organism [1]. Transcriptomics studies usually focus on the levels of expression of these transcripts (i.e., how many copies are found), and this information can be used to define disease biomarkers and predict risks or potential treatment responses [2].

Although different tools for transcriptomics analysis exist, these require bioinformatics expertise, and normally, multiple tools are needed to perform an analysis. Therefore, there is a lack of easy-to-use solutions that allow users with little experience and limited programming knowledge to perform an entire analysis using a single tool.

To fill this gap and provide an easy-to-use solution for transcriptomics analysis, our goal was to assess a typical transcriptomics pipeline and explore tools widely used in this area, and finally develop the transcriptomic analysis modules of OmicSDK. OmicsSDK is a comprehensive software library for the analysis and visualization of omics data. The new modules allow the processing of raw RNA-seq data, differential expression analysis (DEA), functional analysis, and visualization of results using charts.

[1] Corresponding Author: Aurora Sucre, Vicomtech Foundation; E-mail: amsucre@vicomtech.org.

2. Methods

OmicSDK-transcriptomics was developed to ease the analysis and visualization of transcriptomics data and integrate into a single platform all the tools needed for a typical analysis. Initially, a review of literature was followed, then the different requirements were defined and finally the solution was developed, following an iterative process.

2.1. Literature review and definition of requirements

An extensive literature review was performed to define a common processing pipeline for gene expression data, detect which visual analytics approaches could be useful in that context, and define all functional requirements. During this process, no tools were found to perform all the expected steps, but many open-source libraries focusing on specific tasks were detected[2].

Besides the functional requirements defined during the review, it is a key factor that the solution is intuitive and simple enough, so researchers coming from diverse backgrounds and having different programming skills can exploit it.

2.2. OmicSDK-transcriptomics design and development

OmicSDK is a software library composed by three main modules for the ***1) analysis, 2) screening*** and ***3) visualization*** of omics data. For each omics field, different functionalities are included in the three modules for the comprehensive analysis of various datasets. This paper focuses on the sub-modules for transcriptomic analysis, which exploit and combine existing open-source tools to provide a complete pipeline for mRNA and miRNA data.

2.2.1. Analyse-OMICS

This module puts together all functions needed for primary analysis, to process raw sequencing files (FASTQ) and finally obtain total counts or normalized expression values. Different pipelines were implemented for mRNA and miRNA data, using the following tools: ***STAR*** [3] for RNA-seq data alignment, ***CUFFLINKS*** [4] to process RNA-Seq data to obtain the relative abundance of the transcripts, and ***miRge 3.0*** [5], which was designed specifically for miRNA-Seq data analysis and performs all relevant pre-processing steps using other third-party tools such as Cutadapt (adapter trimming), Samtools (managing sequencing files), and Bowtie (short reads alignment).

2.2.2. Screen-OMICS

The second module includes functionalities for secondary and tertiary analysis. Some tools were developed to obtain further insight into the data, and others to compare multiple datasets. The module is based on three R packages: ***TCGAbiolinks*** (differential expression analysis), ***ClusterProfiler*** (functional annotation) & ***GOSim*** (GO terms clustering and classification).

[2] Due to format constraints, we are unable to go into as much depth in this section as expected. For more details on the review, please contact the authors.

2.2.3. Visual-OMICS

The final module encompasses functions for data visualization using plots. Most functions were based on R packages such as ***ggplot2*** (bar plots, heat maps, and dot plots), ***circlize*** (chord diagrams) and ***networkD3*** (sankey diagrams).

2.3. Web development

To make the tool available to researchers with different backgrounds, an intuitive web solution was designed so programming knowledge is not required to use OmicSDK.

The solution is expected to grow iteratively; thus, a modular architecture was chosen to ease the inclusion of new functions without affecting the behaviour of other modules.

The web platform was developed using Angular and different Angular libraries such as Bootstrap to create visual interfaces and D3 to create interactive charts. The Angular front-end is connected to the OmicSDK functionalities through a Flask-based backend.

3. Results

OmicSDK is a software library composed by different modules to analyse several types of omics data, covering all stages of analysis. The functionalities are sorted into three main modules according to the analytical stage where they are used. Regarding transcriptomics analysis, OmicSDK can be used for raw data pre-processing, DEA, functional analysis, and/or data visualization, as shown in **Figure 1**.

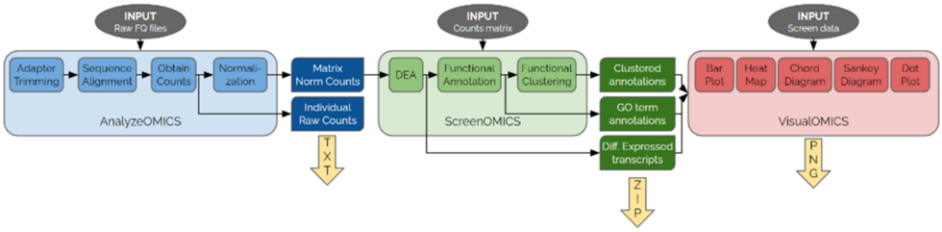


Figure 1. OmicSDK-transcriptomics pipeline: Modules, functionalities, and possible paths to follow.

3.1. Analyse-OMICS

This module allows users to process raw sequencing data to obtain total counts or normalized expression values. It is available for both mRNA and miRNA sequencing data. The following steps are followed to analyse each sample independently: ***1) Adapter trimming*****:** Input FASTQ reads may have adapter sequences at the ends. This initial step ensures only the actual sequence is kept by removing all adapters. ***2) Sequence alignment:*** Given a specific reference genome, all reads are aligned against it to determine how many correspond to each transcript. ***3) Obtain counts:*** During this step, the read counts (number of reads per transcript) are obtained. ***4) Normalization:*** Raw read count values for different samples are not comparable, so a final normalization step is needed. For miRNA data, the RPM (reads per million maps read) value is obtained, but for mRNA data, the RPKM (reads per kilobase per million mapped reads) value is obtained instead, because, in this case, the read length must be considered.

In case the user wants to further analyse and compare a dataset of multiple samples, the module also allows to merge all normalized values in a single matrix or obtain two independent matrices corresponding to different conditions to be compared.

3.2. Screen-OMICS

This module was intended for the comparison of expression data of two well-distinguished groups, as is commonly done in transcriptomics. To do so, three main functionalities were developed: ***1) DEA:*** During differential expression analysis, the normalized counts of two separate groups are statistically analysed to detect significant changes in expression levels between the two groups. For each transcript, a fold change value is obtained, representing how overexpressed or under expressed the transcript is in one group with respect to the other. Also, the statistics, associated p-values and FDR values are obtained, and results can be filtered according to this. ***2) Functional annotation:*** This step is performed to gain further insight into the differentially expressed transcripts and their biological function. The user must provide the list of transcripts, the algorithm for p-value adjustment and which ontologies to consider for annotation (molecular functions, cellular components and/or biological processes). After annotation, a list of GO terms is given. ***3) Functional clustering:*** The last step is to analyse the list of functional annotations to group them according to their similarities. Clustering is performed using the method chosen by the user (options are given).

3.3. Visual-OMICS

The final module includes the tools developed for result visualization, providing users with different graphical elements for a better understanding of the output. The results generated in previous modules are shown in tables and output files, but this module allows plotting these results, specifically using: ***bar plot*** (mean expression values for each transcript, for each group of samples), ***heat map*** (mean expression values, using a colour map), ***chord diagram*** (circular plot showing the relation between genes and their GO functions), ***Sankey diagram*** (2-column plot showing relations between GO terms and transcripts) and ***dot plot*** (mean expression for different group of samples. Distinct groups of genes are plot with different colours). Plot examples are shown in **Figure 2**.

3.4. Web tool and typical pipeline

All functions included in the software library can be used from the command line, so more experienced users can exploit all the benefits. Nevertheless, an angular-flask web solution was also developed, so OmicSDK is available to researchers with different backgrounds, including those with limited programming skills.

Each analytical module was embedded in a different section of the web tool, so they could be used independently. Nevertheless, a communication service was defined, so all elements are connected, and a complete analysis can be performed easily. Users may begin and end their analysis at any desired stage (see **Figure 1**), and all resulting files and plots are downloadable.

A module for interactive visualization was developed exclusively for the web-based tool, where all the described plots can be generated, but interactive utilities that allow users to customize their plots are available.

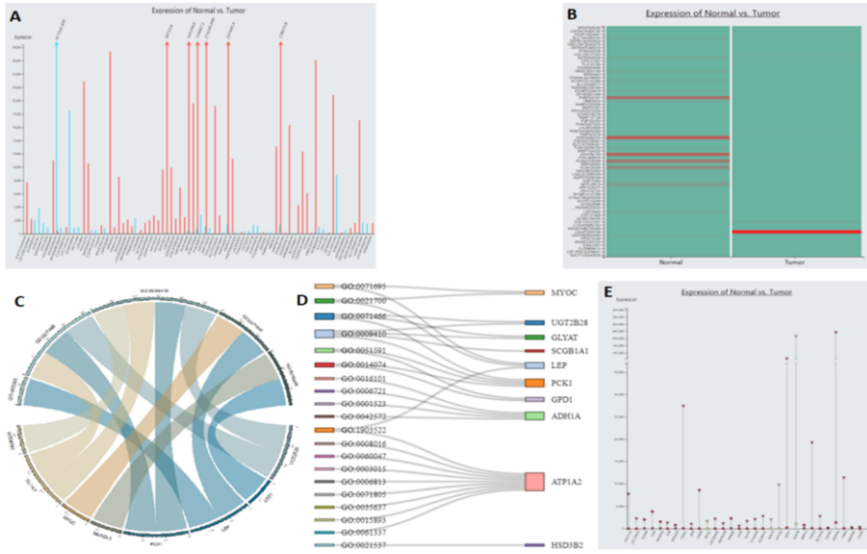


Figure 2. Visual-OMICS plots: A) Bar plot, B) Heat map, C) Chord diagram, D) Sankey plot, E) Dot plot

4. Conclusions

Even if multiple tools are available for RNA-Seq analysis, most of them only implement certain steps of the process or are overly complex to use. In other cases where more comprehensive tools were found, important functionalities were still missing.

OmicSDK-transcripts was designed to cover the main steps of a transcriptomics analysis. Its web implementation, having a user-centred design, should ease the usage for people with different backgrounds. It is expected to be an intuitive and easy-to-use solution, but usability and validation tests are still under design, so its real potential cannot be assured until that stage is reached. However, preliminary tests with subjects with no expertise (students, volunteers with no experience…) returned positive feedback.

The modular nature of the solution allowed its quick integration within the OmicSDK framework, taking advantage of its background capabilities. It will also ease the implementation of other functionalities in the future, when new necessities are detected. Finally, it makes possible to combine the available modules as needed.

References

[1] Ganguly P. Transcription [Internet]. Talking Glossary of Genomic and Genetic Terms. Available from: https://www.genome.gov/genetics-glossary/Transcription

[2] Borràs DM, Janssen B. The use of transcriptomics in clinical applications. Integration of Omics Approaches and Systems Biology for Clinical Applications. 2018;49–66.

[3] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. Star: Ultrafast universal RNA-seq aligner. Bioinformatics. 2012;29(1):15–21.

[4] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene, and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nature Protocols. 2012;7(3):562–78.

[5] Patil AH, Halushka MK. Mirge3.0: A comprehensive microRNA and TRF Sequencing Analysis Pipeline. NAR Genomics and Bioinformatics. 2021;3(3).

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230344

Differential Gene Expression Data Analysis of ASD Using Random Forest

Pragya[a,1], Praveen Kumar GOVARTHAN[a], Kshitij SINHA [b],
Sudip MUKHERJEE[a,1] and Jac Fredo AGASTINOSE RONICKOM[a,1]
[a] *School of Biomedical Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India*
[b] *School of Biochemical Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India*
ORCiD ID: Pragya https://orcid.org/0000-0003-3815-3232

Abstract. Autism spectrum disorder (ASD) is a developmental disability caused by differences in the brain regions. Analysis of differential expression (DE) of transcriptomic data allows for genome-wide analysis of gene expression changes related to ASD. De-novo mutations may play a vital role in ASD, but the list of genes involved is still far from complete. Differentially expressed genes (DEGs) are treated as candidate biomarkers and a small set of DEGs might be identified as biomarkers using either biological knowledge or data-driven approaches like machine learning and statistical analysis. In this study, we employed a machine learning-based approach to identify the differential gene expression between ASD and Typical Development (TD). The gene expression data of 15 ASD and 15 TD were obtained from the NCBI GEO database. Initially, we extracted the data and used a standard pipeline to pre-process the data. Further, Random Forest (RF) was used to discriminate genes between ASD and TD. We identified the top 10 prominent differential genes and compared them with the statistical test results. Our results show that the proposed RF model yields 5-fold cross-validation accuracy, sensitivity and specificity of 96.67%. Further, we obtained precision and F-measure scores of 97.5% and 96.57%, respectively. Moreover, we found 34 unique DEG chromosomal locations having influential contributions in identifying ASD from TD. We have also identified chr3:113322718-113322659 as the most significant contributing chromosomal location in discriminating ASD and TD. Our machine learning-based method of refining DE analysis is promising for finding biomarkers from gene expression profiles and prioritizing DEGs. Moreover, our study reported top 10 gene signatures for ASD may facilitate the development of reliable diagnosis and prognosis biomarkers for screening ASD.

Keywords. Gene expression data, NCBI, Autism Spectrum Disorder, Random Forest, Statistical test

1. Introduction

Autism spectrum disorder (ASD) is a developmental disability characterized by social communication, interaction, and restricted or repetitive behaviors or interests [1,2]. It is caused by environmental and genetic factors. Studies on gene expression can help us to

[1] Corresponding Authors: Pragya, E-mail: pragya.rs.bme22@itbhu.ac.in, Dr. SM, E-mail: sudip.bme@iitbhu.ac.in, E-mail: Dr. JFAR, jack.bme@iitbhu.ac.in.

identify the protein that is primarily responsible for ASD. Moreover, Differential gene expression study helps to understand the biological differences at the genetic level between typical and diseased conditions [3]. Many technologies can capture gene expression from the DNA or RNA such as Microarray DNA, qPCR, and RNAseq [4]. However, these methods have the disadvantage of high cost and time-consuming. Although the list of risk genes implicated by de-novo mutations is growing, it is still very likely far from complete, with an estimated full set of ASD genes ranging from several hundred to more than 30,000. In the search for additional de-novo mutations, sequencing studies continue to be an important approach, but the current sequencing cost is still very high, especially for large samples [5]. As an alternative strategy, advanced analytical approaches like machine learning and statistical methods, which leverage previously implicated genes and prior knowledge, have the potential to enhance risk gene discovery in an efficient and cost-effective manner [6,7,8].

In this study, we have proposed a machine learning-based process pipeline for effectively identifying the candidate chromosomal locations (Hereafter will be referred tas genes for simplicity) in ASD. Random Forest (RF), an ensemble-based classifier, is trained with the gene expression data of ASD and TD from the NCBI GEO database. We have chosen RF, as it's a widely-used and established machine-learning algorithm for classifying datasets with many features. RF is an ensemble method that combines multiple decision trees, handling complex relationships between features and the target variable. It can also estimate feature importance, identifying informative genes for classification. The built machine learning model is validated using 5-fold cross-validation and candidate DEGs responsible for ASD were found.

2. Methods

The processing pipeline adopted in this study is shown in Figure 1. The gene expression data were downloaded from the NCBI GEO database of GSE7329 [9]. The gene expression data provides information about the expression level of the 43,932 genes of 30 samples (ASD=15 and TD=15) [10]. The gene expression data was preprocessed using an algorithm implemented in the R programming language.

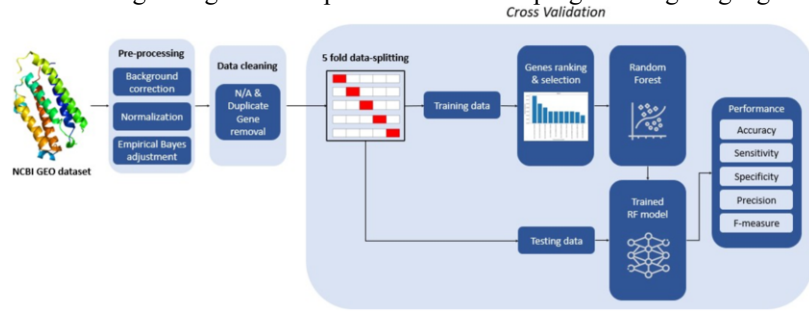


Figure 1. Flowchart of the proposed pipeline

Our study used multiple software packages including GEOquery, LIMMA and Tidyverse to predict disease outcomes based on gene expression data. In addition, the Python programming language package Sci-kit learn 1.0.1 was also used for building and training machine learning models to predict disease outcomes based on gene expression data. Initially, background correction was performed to subtract the background intensity from the foreground intensity for each spot. It attempts to adjust

the data for ambient intensity surrounding each feature. We normalized the gene data and subsequently applied empirical Bayes statistics to differential expression to rank genes in order of evidence for differential expression. The computed linear model fit was then applied to generate the moderated t-statistics, moderated F-statistic, and log-odds of differential expression by empirical Bayes moderation of the standard errors towards a global value. We excluded 2,686 genes from the analysis, as their chromosomal locations were not available in the data set. Furthermore, 2 duplicates of 15 genes and 10 duplicates of 75 genes were observed and we computed the average for these genes. This involved taking the sum of the expression values for each gene and dividing it by the number of times that gene occurred in the dataset. To remove these redundant and missing gene names, we sorted and filtered the data in Microsoft Excel 2016, which led to the exclusion of these genes. Our final analysis included a total gene count of 40,556. We per- formed extensive 5-fold cross-validation to evaluate the performance of the RF machine learning model. Further, we optimized the number of features (Chromosomal locations) for the training model using the feature ranking method (Top 10 Chromosomal locations) of RF. During training, the genes were ranked according to their importance and the top 10 genes were only included in building the model in each fold. We computed the performance metrics like accuracy, sensitivity, specificity, precision and F-measure to evaluate the performance of the classification.

3. Results and Discussions

Figure 2 shows the occurrence of significant genes across 5-fold cross-validation by RF. There are 10 most important genes per fold (50 in total), but due to the overlapping of certain genes in the 5 folds, we ended up with 34 unique genes whose number of occurrences add up to 50. It can be observed that 'chr3:113322718-113322659' (Gene name-germinal centre expressed transcript) was the significant DEG present in every fold. So, we conclude that this gene plays a significant role in ASD. Other genes such as 'chr3:197078850-197078791 and chr4:147533543147533484' occurred in 3 out of 5 folds indicating influential contributions to the classifier. Furthermore, 8 chromosomal locations were present in 2 folds and the rest were present in a single fold.

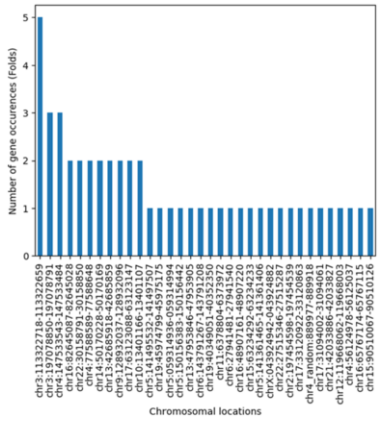


Figure 2. Histogram plot of the number of occurrences concerning chromosomal locations of genes in the 5-fold cross-validation

We obtained 96.67% for the accuracy metrics and the same for the sensitivity and specificity metrics. Similarly, the model also achieved a high precision of 97.5% and an F-measure score of 96.57%. The results showed that our proposed pipeline was able to recognize ASD from TD with high average classification accuracy.

Table 1 depicts the top 10 DEGs identified based on the RF model and statistical method. Our results show that chr3:113322718-113322659 and chrX:72826272-72826213 were the most significant genes by RF and t-test respectively. However, the top 10 genes identified by the RF and t-test were dissimilar.

Table 1. Comparison of top 10 DEGs of ASD identified through RF and t-test methods

S.No.	RF		t-test	
	Chromosomal location	Score	Chromosomal location	p-value
1	chr3:113322718-113322659	0.011	chrX:72826272-72826213	9.66E-33
2	chr3:197078850-197078791	0.006	chr11:61161623-61161682	1.08E-25
3	chr4:147533543-147533484	0.007	chr12:54797596-54797655	4.33E-25
4	chr16:82645087-82645028	0.006	chr6:74284574-74284515	9.88E-25
5	chr22:30158791-30158850	0.008	chr6:33351778-33351946	1.22E-24
6	chr4:77588589-77588648	0.006	chr1:44913389-44913448	1.55E-24
7	chr14:50170228-50170169	0.008	chr2:232403578-232403637	2.6E-24
8	chr13:42685918-42685859	0.007	chr17:34259949-34259890	3.62E-24
9	chr9:128932037-128932096	0.006	chr16:1952138-1952082	4.56E-24
10	chr17:63123088-63123147	0.006	chr7:5340087-5340028	5.17E-24

4. Limitations and Future work

Our process pipeline has produced a high classification accuracy of 96.7% to discriminate between ASD and TD. However, it has a few limitations. We ranked and selected the top 10 genes but failed to find similar genes between RF and t-test analysis. We can select the top 20, 30, or 40 genes and then re-validate the results. We can also use other preprocessing pipelines, like low-level preprocessing or high-level preprocessing methods [11]. In addition, we have considered only a single gene expression dataset and can use more datasets for improved performance. We built the models using machine learning classifiers but never attempted deep learning algorithms. We used only RF; however, we can use other classifiers such as support vector machine, logistic regression, XGBoost etc.

5. Conclusions

We have proposed a machine learning framework for identifying the genes responsible for causing ASD in an individual. We achieved an accuracy of 96.67% and the top performing gene was the germinal center expressed transcript 2 (chromosomal location chr3:113322718-113322659) by the RF algorithm. However, with the statistical analysis, we found the chromosomal location chrX:72826272-72826213 was responsible for ASD. The chromosomal locations were found different in both approaches. However, the RF model produced high classification accuracy. Our study

shows the possibility of utilizing the proposed model in a potential application for screening ASD and TD in a clinical environment.

References

[1] Ansel A, Rosenzweig JP, Zisman PD, Melamed M, Gesundheit B. Variation in gene expression in autism spectrum disorders: an extensive review of transcriptomic studies. Frontiers in neuroscience. 2017 Jan 5;10:601.

[2] Ronicko JF, Thomas J, Thangavel P, Koneru V, Langs G, Dauwels J. Diagnostic classification of autism using resting-state fMRI data improves with full correlation functional brain connectivity compared to partial correlation. Journal of Neuroscience Methods. 2020 Nov 1;345:108884.

[3] Rylaarsdam L, Guemez-Gamboa A. Genetic causes and modifiers of autism spectrum disorder. Frontiers in cellular neuroscience. 2019:385.

[4] Nagy A´ , La´nczky A, Menyha´rt O, Gyo˝rffy B. Validation of miRNA prognostic power in hepatocellular

[5] Carcinoma using expression data of independent datasets. Scientific reports. 2018 Jun 15;8(1):1-9.

[6] Lin Y, Afshar S, Rajadhyaksha AM, Potash JB, Han S. A machine learning approach to predicting autism risk genes: Validation of known genes and discovery of new candidates. Frontiers in genetics. 2020 Sep 10;11:500064.

[7] Asif M, Martiniano HF, Vicente AM, Couto FM. Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology. PloS one. 2018 Dec 10;13(12):e0208626.

[8] Go¨k M. A novel machine learning model to predict autism spectrum disorders risk gene. Neural Computing and Applications. 2019 Oct;31(10):6711-7.

[9] Brueggeman L, Koomar T, Michaelson JJ. Forecasting risk gene discovery in autism with machine learning and genome-scale data. Scientific reports. 2020 Mar 12;10(1):1-1.

[10] Agastheeswaramoorthy K, Sevilimedu A. Drug REpurposing using AI/ML tools-for Rare Diseases (DREAM-RD): A case study with Fragile X Syndrome (FXS). bioRxiv. 2020 Jan 1.

[11] Nishimura Y, Martin CL, Vazquez-Lopez A, Spence SJ, Alvarez-Retuerto AI, Sigman M, Steindler C, Pellegrini S, Schanen NC, Warren ST, Geschwind DH. Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. Human molecular genetics. 2007 Jul 15;16(14):1682-98.

[12] Dziuda DM. Data mining for genomics and proteomics: analysis of gene and protein expression data. John Wiley & Sons; 2010 Jul 16.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230345

Towards a Digital Twin in Human Brain: Brain Tumor Detection Using K-Means

Anastasios Loukas SARRIS [a] Efstathios SIDIROPOULOS [a]
Evangelos PARASKEVOPOULOS [b] and Panagiotis BAMIDIS [b]

[a] *Anaptixiaki Meletitiki Voriou Ellados*
[b] *Lab of Medical Physics and Digital Innovation*

Abstract. Digital Twins come to revolutionize the ongoing procedures of healthcare industry, with their ability to stimulate and predict patients' diagnosis and treatment. In this paper a K-means based brain tumor detection algorithm and its 3D modelling design, both derived from MRI scans, are presented towards to the creation of the digital twin.

Keywords. Digital Twin, Brain Tumor, Clustering, K-means

1. Introduction

At the dawn of the 4th industrial revolution, Digital Twin (DT) comes to be at the center of study. Digital Twin technology is the third trending technology for 2020[1]. A Digital Twin (DT) is a virtual replica of a physical object or a system and reflects its properties. A Digital Twin environment provides a rapid analysis and real-time decisions made through accurate analytics, running multiple processes at once. This is made possible with the acquisition of real time data, via technologies such as sensors. 3D Scanning and Machine Learning techniques are applied on the collected data in order to be used on the replica[2]. Nowadays, DTs have an impact on many different fields. Some of them are industry, manufacture, smart cities and healthcare. More specifically, in healthcare, the digital twin technology comes to revolutionize the ongoing procedures of patients' treatment and prescriptions. The digital twins are facilitated with the continuously increased volume of the available data from the Internet of Things (IoT) environment and advanced data analysis, all of which aim at a more personal examination of each patient, known as precision medicine. The most common use of a digital twin in healthcare is to replicate a certain body organ in order to monitor it. The most replicated one is the human heart, but a challenging one that has started to be studied is the human brain[3]. Another medial case where Digital Twins prove themselves useful is in the maintenance of the devices used, such as an MRI scanner[13]. Brain tumors are a mass of growing cells in the brain. They assort in benign, the non-cancerous ones, and malignant tumors[6]. They have high mortality rate, which can be seen from the fact that 35.6% of people diagnosed with brain tumor live five years after the tumor detection[9].

2. Related Work

Image Segmentation is the technique to partition the digital image in separate groups. All of the segments in a group have similar pixel features with each other[8]. Mustaqeem A et al. proposed an algorithm which uses different filters to denoise the gray level images. They used watershed and threshold segmentation to segment the image on its different parts. The separation of the tumor from the rest of the MRI was achieved using morphological operators [4]. Masood M et al. used Mask RCNN to classify either tumorous or non tumorous images following a median filter noise removal preprocessing [5]. SVM classification and histogram based segmentation was implemented by Chinnu A. in MRI brain images after noise removal and applying a skull stripping extracted the useful features to feed the classifier [6]. Hossam MM et al. used K- Means with component labelling to cluster the 2D brain slices. Next they rendered the object by using some of its features, they visualized in 3D the final outcome [7].

As it was mentioned the DT in healthcare are used to replicate a certain organ. Pajaziti E et al. created a Virtual Reality (VR) system of human heart from six different patients. The region of interest from each patient was chosen manually and the 3D model was created. It aimed to create an educational VR model of the human heart[14]. Capellini K et al. proposed a novel semi-automatic framework to prepare the right anterior mini-thoracotomy surgery from the 3D model of the aorta and the anterior rib cage (aRC). They used region growing segmentation algorithm to create the aorta model from CT images. For the visualization of the aRC they traced the first five ribs and the sternum in axial and sagittal views. These parts were transformed in polylines and were joined together to create the second part of the 3D model[15].Kardampiki E et al. proposed a framework that combines computational fluid dynamics alongside 3D model of the aorta in order to create a DT pipeline that evaluates the parameters of Modified Blalock Taussig Shunt (MBTS). The ascending/descending aorta sections, aortic arch, epiaortic vessels, the pulmonary branch and the valve plane were segmented so the model was created. Based on that the MBTS part was recognized and with the fluid equations that were created the parameters of the surgery were evaluated [16]. Poletti G et al. created coronary artery DTs of different patients who undergo stent implantation surgery. They used 2D characteristics from different views to combine the information table of the 3D model. Then with the data derived from the OCT scans and the table they constructed the 3D model [17].

3. Methods

3.1. Proposed Framework

The proposed framework is a data driven scenario [12]. The model is constantly fed with updated data that concern the patient. The actual MRI scans of the subject were used in order to create and monitor the patient's brain replica. This is followed by a tremendous amount of data taken from its medical history and real time data from sensors such as mobile EEG technologies etc. The EEG recordings of the patient are used in order to collect useful data, along with the ones provided from the IoT environment, aiming at strengthening the decision making capacity of the twin. Based on that the implementation

of a brain tumor detection algorithm is a first step towards the decision making capability of the system.

3.2. Tumor Detection Methodology

The implementation of brain tumor detecting algorithm is based on brightness contrast analysis in MRI scans. It is meant that the tumor has a different, usually more intense, brightness level. This is where a clustering technique was required. K-means is the simplest clustering method that is based on the concept of minimizing the squared distance from the cluster center [7]. The process that was followed is described as a colorized clustering. The input is the 3D MRI image itself where the third dimension is the RGB combination of each pixel. Every cluster is created after the implementation of K-means representing a unique, same-value combination of RGB components. The optimal number of clusters is decided in advance based on the knee rule fro clustering each one of the scans. The Within cluster Sum of Squares or Sum of Squared Errors (SSE) metric shows the euclidean distance sum of all the cluster elements from the cluster centers.

$$SSE = \sum_{n=1}^{cluster\_elements} dist^2(n, cluster\_center) \tag{1}$$

Cluster centers are not a Cartesian combination of points but they are a list containing the RGB components of the cluster color, which defines the RGB combination of the whole cluster. From the clustered image was selected the class which had the highest, value wise, color combination. After the isolation of the pixels a circular mask, with dynamic radius is created. The initial radius value and the center pixels' combination of the mask are specified. The mask checks which pixels within the radius have the same color values as the central one and computes their area within the mask. This process is implemented for each one of the isolated pixels. If the area is over a threshold value then it is considered to be the tumor. Due to the fact that the data set consists of many diverse MRI images the radius adapts itself in any given subject so if the procedure is unable to detect the tumor the radius changes to a smaller value and the masking check restarts.

3.3. 3D Modeling

The right branch of the Figure 1 shows the process to 3D model. The whole design is accomplished with the open source computer-aided design (CAD) software 3D Slicer, using its built-in modules and some extensions provided from the software itself. An automated procedure is aimed to create a 3D model for each subject, on condition that the scans are appropriate.

4. Results

4.1. Data Acquisition

To evaluate the performance of the methodology described previously we used different MRI images taken from different MRI scanners and different conditions. Our data set consisted of 247 MRI images obtained from 11 different Radiopaedia cases[10] and

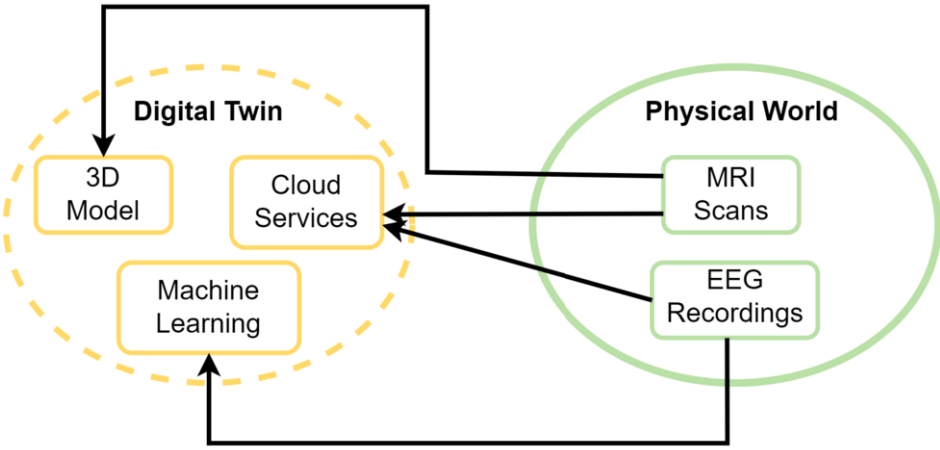


Figure 1. MRI scans processing towards the brain 3D Model and the tumor detection

Brain MRI Images for Brain Tumor Detection by Kaggle[11]. The 59.91% of them referred to patients who suffered from brain tumor and 40.09% referred to healthy patients. The current image format is JPG.

4.2. Experimental Results

The tumor detection algorithm was applied on all 248 scans with different radii varying from 30 to 100. In the non tumorous ones the radius was limited between 30 and 75 and it was observed that the greater the value of the radius is the greater the accuracy score was. The accuracy is defined as the percentage of the images that the algorithm did not find any tumor to the total number of images.

$$Accuracy = \frac{Number\_of\_non\_Tumorous\_Images}{Total\_Number\_of\_Images} \tag{2}$$

This process was made in order to make sure that the algorithm is capable of identifying MRI images of healthy patients from them with tumors. So based on Eq. 2 the accuracy was varying from 0.85% to 0.96% (Mean: 0.92, std: 0.04), which for sure is not realistic due to the limited number of healthy patients' samples that we possessed. As it is shown from the first two steps of Figure 2 it is easily seen that the tumor has been successfully located. Then the 3D model is created by the combination of tumor and brain part visualizations.

5. Future Work

Our future work extends in generalizing the algorithm so it does not only identify tumors as the brightest object of the image but in all cases and remove outliers. In order to achieve that the process will be tested on NIFTI images. Another aspect that will be investigated is the relation of the EEG when brain tumors are present, aiming on more secure decision by the twin. Finally to achieve the ultimate goal of the brain 3D model an attempt of creating the tumor 3D directly after its detection will be implemented.

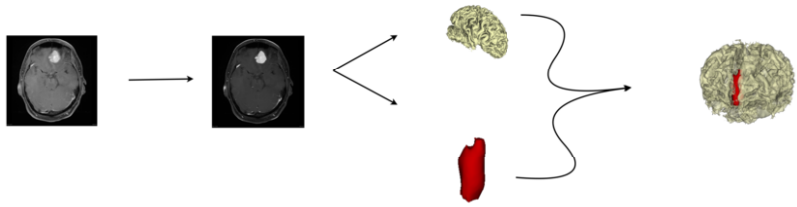


Figure 2. The implementation result. Starting from an MRI, the tumor is detected and then the 3D model of brain and the tumor is constructed .

References

[1] Elayan H, Aloqaily M, Guizani M. Digital twin for intelligent context-aware IoT healthcare systems. IEEE Internet of Things Journal. 2021 Jan 12;8(23):16749-57, doi: 10.1109/JIOT.2021.3051158

[2] Liu M, Fang S, Dong H, Xu C. Review of digital twin about concepts, technologies, and industrial applications. Journal of Manufacturing Systems. 2021 Jan 1;58:346-61, doi: 10.1016/j.jmsy.2020.06.017

[3] G. Browne, "The quest to make a digital replica of your brain."[Internet] 2022. Available from: https://www.wired.com/story/the-quest-to-make-a-digital-replica-of-your-brain/

[4] Mustaqeem A, Javed A, Fatima T. An efficient brain tumor detection algorithm using watershed & thresholding based segmentation. International Journal of Image, Graphics and Signal Processing. 2012 Sep 1;4(10):34, doi: 10.5815/ijigsp.2012.10.05

[5] Masood M, Nazir T, Nawaz M, Javed A, Iqbal M, Mehmood A. Brain tumor localization and segmentation using mask RCNN. Frontiers Comput. Sci.. 2021 Dec 1;15(6):156338, doi: 10.1007/s11704-020-0105-y

[6] Chinnu A. MRI brain tumor classification using SVM and histogram based image segmentation. International Journal of Computer Science and Information Technologies. 2015;6(2):1505-8.

[7] Hossam MM, Hassanien AE, Shoman M. 3D brain tumor segmentation scheme using K-mean clustering and connected component labeling algorithms. In2010 10th International Conference on Intelligent Systems Design and Applications 2010 Nov (pp. 320-324). IEEE, doi: 10.1109/ISDA.2010.5687244

[8] The Mathworks Inc., "What Is Image Segmentation?"[Internet] 2022. Available from: https://ch.mathworks.com/discovery/image-segmentation.html

[9] Cancer Treatment Centers of America - CTCA,"Brain cancer grades"[Internet] 2022. Available from: https://www.cancercenter.com/cancer-types/brain-cancer/grades

[10] ,"Radiopaedia" [Internet] 2022. Available from: https://radiopaedia.org/?lang=us

[11] Navoneel Chakrabarty, "Brain MRI Images for Brain Tumor Detection" [Internet] 2018. Available from: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection

[12] Sarris A. L., Sidiropoulos E., Paraskevopoulos E., Bamidis P. Digital Twins in Human Brain: Opportunities and Challenges. Poster presented at: Society of Applied Neuroscience. SAN2022. 2022 September 15-17; Thessaloniki

[13] Pesapane F, Rotili A, Penco S, Nicosia L, Cassano E. Digital Twins in Radiology. Journal of Clinical Medicine. 2022 Nov 4;11(21):6553.

[14] Pajaziti E, Schievano S, Sauvage E, Cook A, Capelli C. Investigating the Feasibility of Virtual Reality (VR) for Teaching Cardiac Morphology. Electronics. 2021 Aug 6;10(16):1889.

[15] Capellini K, Positano V, Murzi M, Farneti PA, Concistrè G, Landini L, Celi S. A Decision-Support Informatics Platform for Minimally Invasive Aortic Valve Replacement. Electronics. 2022 Jun 17;11(12):1902.

[16] Kardampiki E, Vignali E, Haxhiademi D, Federici D, Ferrante E, Porziani S, Chiappa A, Groth C, Cioffi M, Biancolini ME, Costa E. The Hemodynamic Effect of Modified Blalock–Taussig Shunt Morphologies: A Computational Analysis Based on Reduced Order Modeling. Electronics. 2022 Jun 21;11(13):1930.

[17] Poletti G, Antonini L, Mandelli L, Tsompou P, Karanasiou GS, Papafaklis MI, Michalis LK, Fotiadis DI, Petrini L, Pennati G. Towards a Digital Twin of Coronary Stenting: A Suitable and Validated Image-Based Approach for Mimicking Patient-Specific Coronary Arteries. Electronics. 2022 Feb 8;11(3):502.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230346

Challenges of Estimating Global Feature Importance in Real-World Health Care Data

Aniek F. MARKUS[a,12], Egill A. FRIDGEIRSSON[a], Jan A. KORS[a], Katia M.C. VERHAMME[a] and Peter R. RIJNBEEK[a]
[a] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands*

Abstract. Feature importance is often used to explain clinical prediction models. In this work, we examine three challenges using experiments with electronic health record data: computational feasibility, choosing between methods, and interpretation of the resulting explanation. This work aims to create awareness of the disagreement between feature importance methods and underscores the need for guidance to practitioners how to deal with these discrepancies.

Keywords. Prediction modelling, Explainable AI, Variable importance, Shapley values, Evaluating explanations

1. Introduction

Personalized medicine aims to provide treatment and prevention tailored to individual patients. Machine learning (ML) models can help with personalized risk prediction based on individuals' characteristics to pursue this goal. When implementing prediction models in practice, explanations can be useful to validate model behavior and/or to create a shared meaning of the decision-making process [1].

Feature importance is often used to explain ML models and identified as useful explanation by clinicians [2]. A higher score implies a higher importance of the specific feature, i.e. a larger impact on the model predictions ('How does the output rely on a variable?') or model performance ('How much is the loss function reduced?'). In the literature, many methods have been proposed to compute feature importance. In this work we focus on model-agnostic methods (i.e. suitable to explain any kind of ML model) to compute global feature importance (i.e. explaining the model as a whole).

In practice, there are several challenges when aiming to estimate feature importance for prediction models developed using electronic health record (EHR) data: I) computational feasibility, II) choosing between methods, and III) interpretation of the resulting explanation. These challenges are not unique to EHR data, but might be magnified due to the large size, high-dimensionality, and sparsity of the data. In the following paragraphs we discuss the three challenges in more detail:

[1] Corresponding Author: A.F. Markus, E-mail: a.markus@erasmusmc.nl.

[2] This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

I) First, not all methods are computationally feasible for big data. Some feature importance methods require examining all possible combinations of K features or rely on conditional distributions which are unavailable in practice and difficult to estimate. Although there is already work dealing with various ways of approximations (e.g. for Shapley values), formal evaluation of methods (and their approximations) is often lacking and otherwise focused on relatively low-dimensional data.

II) Second, there is a need to choose between methods, but guidance on which method is best to use is lacking. In existing studies, the reason for preferring the chosen feature importance method over alternatives is rarely motivated. However, there is increased awareness that different feature importance methods do not align on the generated explanations (e.g. ranking of top features). This makes the 'interchangeable use' of methods problematic. In recent work, Krishna et al. [3] formalized this as the disagreement problem and analyzed how users deal with this problem in practice.

III) Finally, feature importance explanations might not be in line with (user) expectations. Hase et al. [4] state that explanations are socially misaligned when they convey a different kind of information than what users expect. As an example, they mention some unexpected factors (e.g. model seed, hyperparameters) that might influence the resulting explanations more than expected factors (e.g. the data). However, there are also large differences in how feature importance methods work (e.g. different definitions and/or assumptions), which can formally explain variation in explanations. Resulting explanations miss this nuance and are interpreted similarly by end-users. The general lack of consensus on a definition for feature importance makes it impossible to systematically evaluate whether the selected subset of features is truly important.

In this work, we examine these challenges using experiments with real-world health data when estimating feature importance for a model predicting hospital readmission within 30 days.

2. Methods

2.1. Real-world data

We developed a prediction model on the Dutch Integrated Primary Care Information (IPCI) database [5] to answer the following question: "Among adult patients discharged from the hospital (target population), which patients will be readmitted (outcome) within 30 days (time-at-risk) after the visit?". The IPCI database has been mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), which enables standardized extraction and analysis of health care data. This study was approved by the IPCI Governance Board (number 09/2020).

2.2. Model development

We selected a random sample of N = 100,000 patients of which 75% was used for model development ('training set'). The remaining 25% of patients ('test set') was used for validation. We trained prediction models using the python library `sklearn` using logistic regression with L2-regularization (C=0.1) on the training set. Candidate covariates included sex, age (in 5-year groups), and binary variables indicating the presence or absence of recorded conditions and drugs (measured 30 days, 1 year, and any time prior

to index). To obtain a dataset with dimensionality K, we selected the K features with the strongest relation to the outcome based on the Pearson correlation coefficient.

2.3. Feature importance methods

We investigated the following feature importance methods:

- **Permutation feature importance (FI)**: measures the decrease in model predictive performance after random shuffling the values of a certain feature. We measured model performance using the area under the receiver operator curve (AUC) and balanced accuracy (BA). This method has the advantage that it is fast because it does not require retraining of the model, but has the disadvantage that it might extrapolate to unlikely data points.
- **Shapley Additive exPlanation (SHAP) value**: measures the marginal contribution of features, averaged over all orderings in which the subset of features can be constructed. The resulting explanations are considered to result in a 'fair' allocation because they satisfy five desirable properties (efficiency, symmetry, dummy, monotonicity, and linearity). However, depending on the estimation strategy the method can still extrapolate to unlikely data points in case of correlated features. The computation time of exact Shapley values is an NP-hard problem, therefore we studied two approximations:
  - **KernelSHAP**: uses a weighted linear regression to estimate local SHAP values [6], we implemented this using the python library `shap` (with nsamples = 10*num_features + 2048 and l1_reg = 'num_features(10)').
  - **SAGE**: a sampling-based approximation to compute global SHAP values [7], we implemented this using the python library `sage-importance` (with MarginalImputer() and n_permutations = 1000*num_features).

2.4. Experiments

I. Computational feasibility: we investigated which feature importance methods are able to deal with the high dimensionality of EHR data. We measured the computation time when calculating feature importance using each method across different data dimensionalities K = [20, 50, 100]. For Shapley values we evaluated using 500 and 1000 background samples. The experiments were run using 16 cores of an Intel® Xeon® CPU E5 v4.

II. Choosing between methods: we investigated to what extent different feature importance methods result in different rankings of features for EHR data. We investigated the top-10 ranked features as users typically focus only on the most important features.

III. Interpretation of the resulting explanation: we investigated alignment with user expectations. We argue that users expect feature importance for additive models to be in line with the size of the model coefficients. We measured alignment of the feature importance methods with model coefficients, comparing both the ranking (using top-5 overlap, top-5 sign agreement, and Kendall's tau) and normalized values (using mean absolute error).

3. Results

3.1. Experiment I: computational feasibility

Table 1 indicates the required computation times for each of the methods. This shows that the computation times to explain a prediction model quickly explode. Permutation methods are very fast, but the approximations of Shapley values take up to multiple hours. These results suggest that it is critical to further improve the speed of Shapley methods as the explanation times might be a hurdle for model explanation in practice.

Table 1. Computation time (in minutes) for various FI methods.

	Permutation FI		**KernelSHAP**		**SAGE**	
	AUC	BA	500 samples	1000 samples	500 samples	1000 samples
K=20	0,0	0,0	23,7	93,4	1,2	2,5
K=50	0,1	0,1	34,6	129,8	16,6	33,6
K=100	0,2	0,3	52,5	193,8	129,2	260,5

3.2. Experiment II: choosing between methods

The prediction models for K = [20, 50, 100] resulted in performance (AUC) of 0.66, 0.66, and 0.67, respectively, on the test set. Figure 1 presents the most important features identified by each method for the prediction model for K=50. This shows significant differences in the explanations across methods. Not only is there a mismatch in the top features, but also the ordering of features and the direction of effect differ (e.g. X28). Model coefficients, permutation FI, and SAGE roughly identify the same set of important features. However, KernelSHAP leads to a very different set of features. SAGE can additionally capture the direction of effect as this does not capture the change in model error (as permutation FI), but the difference between the actual and average prediction.

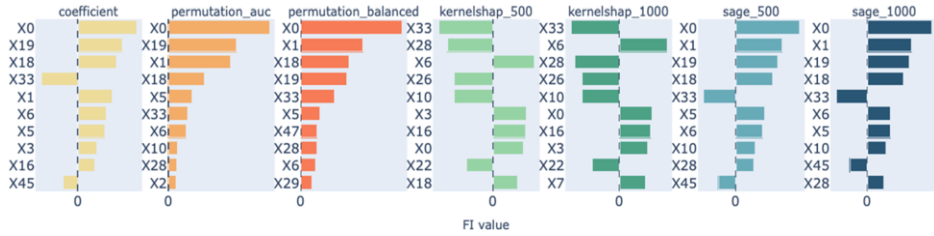


Figure 1. Top-10 ranked features in the prediction model according to different FI methods (K = 50).

3.3. Experiment III: interpretation of the resulting explanation

Figure 2 shows the alignment of feature importance methods with model coefficients measured using three rank-based metrics and one value-based metric. The agreement between model coefficients is highest with SAGE across all metrics, as can be seen from the figure because Top-5/Sign agreement are both 1 (indicating perfect agreement) and the MAE is low (indicating the normalized values are close). These metrics make the discrepancies between model coefficients and KernelSHAP very clear for the top features (Top-5/Sign agreement are both 0.2), but also show that the overall produced ranking (Kendall's tau) and values (MAE) are better than for permutation FI.

Figure 2. Agreement of FI methods with model coefficients. The metrics are scaled between 0-1, with values closer to 1 indicating more agreement.

4. Discussion and Conclusion

It is often important to understand which features are most important for a given prediction model. We have shown there are challenges in different phases of creating such explanations for EHR data, even for a simple classifier. First, the computation times for state-of-the-art feature importance methods are significant, which may be a hurdle to implementation in practice. Second, different feature importance methods often result in different explanations, hence it is important to make an informed choice between methods. Third, even though these observed differences are not always unexpected, e.g. permutation FI/SAGE explains model performance and KernelSHAP explains model predictions, it remains a challenge to communicate these differences to end users.

We only investigated one type of classifier and one prediction task. Results may vary depending on the studied example, but the main findings (e.g. disagreement between methods) have been found in other studies as well (e.g. [3]). Moreover, when investigating other classifiers such as tree-based and/or deep learning methods we expect the problems will only be larger due to the non-linearity of these methods.

This work aims to create awareness of the disagreement between feature importance methods and underscores the need for guidance to practitioners how to deal with the discrepancies between feature importance methods. For this, we argue it is important to make explicit what we mean with feature importance (also for non-linear models) and which goal we aim to fulfill (e.g. to understand model decisions or to improve the model), as this can guide how methods should be formally evaluated and selected.

References

[1] Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. doi: 10.1016/j.jbi.2020.103655.

[2] Tonekaboni S, et al. What clinicians want: Contextualizing explainable machine learning for clinical end use. Proceedings of Machine Learning research. 2019:1-21.

[3] Krishna S, Han T, Gu A, Pombra J, Jabbari S, Wu S, et al. The disagreement problem in explainable machine learning: A practitioner's perspective. arXiv preprint arXiv:220201602. 2022.

[4] Hase P, Xie H, Bansal M. The out-of-distribution problem in explainability and search methods for feature importance explanations. Adv Neural Inf Process Syst. 2021;34.

[5] de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, et al. Data resource profile: The integrated primary care information (IPCI) database, the Netherlands. Int J Epidemiol. 2022. doi: 10.1093/ije/dyac026

[6] Lundberg SM, Lee S-I, editors. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30; 2017.

[7] Covert I, Lundberg S, Lee S-I. Understanding global feature contributions with additive importance measures. Adv Neural Inf Process Syst. 2020;33.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230347

How Good Is ChatGPT for Medication Evidence Synthesis?

Hao LIU[a], Yifan PENG[b] and Chunhua WENG[a,1]
[a]*Dept. of Biomedical Informatics, Columbia University, New York, NY, USA*
[b]*Dept. of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA*

Abstract. With its seeming competence to mimic human responses, ChatGPT, an emerging AI-powered chatbot, has spurred great interest. This study aims to explore the role of ChatGPT in synthesizing medication literature and compare it with a hybrid summarization system. We tested ten medications' effectiveness with reference to their definitions and descriptions extracted from DrugBank. ChatGPT could generate coherent summaries that are not backed by evidence. In contrast, our approach can provide a highly structured and concise synthesis of related evidence, but the resulting summary is not as fluent and convincing as ChatGPT. Therefore, we recommend integrating both techniques to achieve the best performance.

Keywords. Summarization, ChatGPT, Natural Language Processing

1. Introduction

Multi-document summarization is a technique used to create a summary from multiple documents or textual sources. It has long been used for evidence synthesis [1]. Recently, the large language models, exemplified by the recent sensational ChatGPT [2], have taken advantage of pre-training hundreds of billions of parameters on a large corpus of text and have achieved impressive performance in many NLP tasks such as QA or summarization. For example, when asked, "What is the most effective treatment for patients with advanced lung cancer?" ChatGPT can retrieve relevant articles from its training data, summarize the key findings, and provide a summary: "*The most effective treatment for advanced lung cancer can depend on several factors, including the type and stage of cancer, the patient's overall health and medical history, as well as their preferences and goals of care. 1. Chemotherapy … 2. Targeted therapy, … 3. Immunotherapy … 4, Radiation therapy… 5. Surgery…".* The variety of potential text summarization applications and the complexity of model training make ChatGPT particularly appealing for the medical literature synthesis [3]. In this study, we present the first comparative study of ChatGPT and a hybrid multi-documentation summarization method on medication evidence synthesis. We use randomized controlled trials (RCTs) abstracts on PubMed as lengthy documents because they are considered the most reliable source for robust medical evidence for clinical question answering [4] and evidence-based medicine [5]. Our findings suggest the potential and limitations of

[1] Corresponding Author: Department of Biomedical Informatics, Columbia University, 622 West 168th Street, PH-20, Room 407, New York, NY 10032, USA. E-mail: chunhua@columbia.edu.

using ChatGPT for medication evidence synthesis and shed light on the future directions for developing evidence synthesis systems.

2. Method

2.1. Clinical questions on drug effectiveness

We consulted with clinicians and collected ten questions (Table 1) on the effectiveness of drugs for COVID-19, Alzheimer's disease, kidney diseases, and rheumatic diseases.

Table 1. Ten clinical questions on drug effectiveness.

Topic	Questions
COVID-19	Is Hydroxychloroquine (HCQ) effective for treating COVID-19 patients? Is Remdesivir effective for treating COVID-19 patients? Is Tocilizumab effective for treating patients with COVID-19?
Alzheimer's disease	Is Galantamine effective for improving cognitive function for Alzheimer's disease patients? Is Donepezil effective for improving cognitive function for Alzheimer's disease patients?
Rheumatic diseases	Is Tofacitinib effective for treating patients diagnosed with rheumatoid arthritis? Is Belimumab effective for inducing renal remission in patients diagnosed with Systemic Lupus Erythematosus (SLE)? Is Rituximab effective for inducing clinical remission in patients diagnosed with Antineutrophil cytoplasmic antibody (ANCA) vasculitis? Is Cyclophosphamide effective for inducing clinical remission in patients diagnosed with Antineutrophil cytoplasmic antibody (ANCA) vasculitis?
Kidney disease	Is Tolvaptan effective for treating Autosomal dominant polycystic kidney disease (ADPKD)?

2.2. The proposed document summarization system

Our system consists of five modules (Figure 1). **(1) Document collection.** We retrieved PubMed abstracts for 189,648 clinical trial publications by identifying PMIDs labeled "Randomized Controlled Trial" between January 2010 and October 2021. Metadata, such as title, abstract, and metadata, were extracted as JSON files. **(2) Document retrieval.** We employed a two-step approach to retrieve the top-*k* relevant PubMed abstracts for each clinical question, similar to the method used in VERT5ERINI [6]. First, we treated the input question as a "bag of words" and retrieved a set of *n* candidate articles using the BM25 scoring function [7]. Then, we employed an advanced encoder-decoder model (T5 [8]) to estimate the relevance of each candidate article to the question. The top-k articles, ranked by their estimated relevance score, were returned. **(3) Sentence extraction**. Here, we selected the most relevant sentence from each abstract relevant to the question. We used the same T5 model to rank sentences in a study and

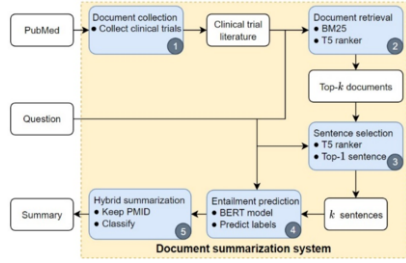


Figure 1. Overview of the document summarization

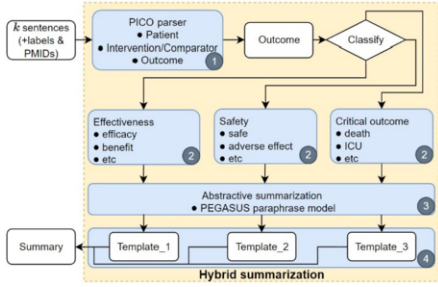


Figure 2. The overview of the hybrid summarization module

chose the top-one sentence as the rationale sentence. **(4) Entailment prediction.** We developed an entailment prediction module that predicts the relationship between the question and the highly ranked sentences. We treated it as a multiclass classification problem, where the output indicates whether a given rationale sentence Supports, (is) Neutral, or Refutes the question. We used a pre-trained model PubMedBERT [9], fine-tuned on entailment datasets. **(5) Hybrid summarization.** This module automatically generates a summary based on all selected sentences from extracted studies via a combination of four components: PICO (Population, Intervention Comparison, and Outcome) parser, study classification, abstractive summarizations, and template formation (Figure 2). The selected sentences were first parsed into PICO entities which are widely-used knowledge representations for clinical questions posed in the natural language [10]. Based on the Outcome entity identified, each sentence is classified into the Effectiveness, Safety, and Critical outcome categories, respectively. Then an abstractive summarization model PEGASUS was used to generate a partial summary of sentences from each category [11]. The three partial summaries were then organized and formatted in a coherent and readable summary following pre-defined templates.

2.3. The comparison study

To get ChatGPT's report on drug effectiveness, each question in Table 1 was queried on the ChatGPT (https://chat.openai.com/chat, version Jan 9, 2023) with responses recorded. In addition, the same questions were also queried against a newly proposed document summarization system (Section 2.2).

The generated summaries are compared to the reference texts manually extracted from DrugBank [12] (Figure 3). DrugBank is a widely adopted, free-access, public website that provides information on drugs, including drugs' chemical, pharmacological, and pharmaceutical properties. First, we selected a drug's overall summary and its pharmacological indication as the description of a drug's effectiveness (i.e., reference text). Then, we paired the reference text with the summaries generated by ChatGPT or our proposed method. Finally, we calculated Rouge [13], BLEU [14], and Levenshtein Distance [15] scores. These three metrics are commonly used to evaluate the automatic summarization of texts against a set of reference texts.

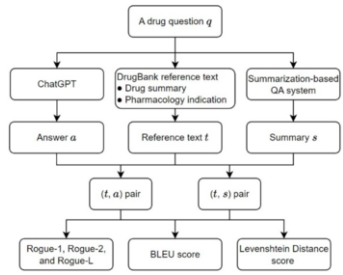


Figure 3. Evaluation process

3. Results and Discussion

Table 2 shows that ChatGPT achieved consistently higher Rouge-1, Rouge-2, and Rouge-L scores than the summarization-based method. The two methods achieved similar averaged BLUE scores (0.041 vs. 0.038). Our method outperformed ChatGPT with a lower averaged Levenshtein distance (1.42 vs. 1.45). Table 3 shows the description of the drug Galantamine in DrugBank and summaries generated by ChatGPT and our method. ChatGPT only described the function of Galantamine and its target disease and warned users of its usage. In contrast, our method was able to generate reports with references to the findings in each study identified in the clinical publication database. This work employed a multi-document summarization system that summarizes

the drug's effectiveness. The findings in such a summary can be linked to research publications or data from clinical trial publications. We also tested the capability of ChatGPT to generate a summary of the effectiveness of drugs. It is important to note that the report generated by ChatGPT was based on the data provided during its training, without differentiating the validity and reliability of the data source.

Table 2. Comparison between ChatGPT and our method.

		ChatGPT	Our method
Rouge-1	Precision	0.333±0.13	0.242±0.09
	Recall	0.207±0.08	0.160±0.06
	F1	0.236±0.07	0.180±0.04
Rogue-2	Precision	0.135±0.09	0.030±0.02
	Recall	0.070±0.05	0.019±0.02
	F1	0.082±0.04	0.021±0.01
Rogue-L	Precision	0.317±0.13	0.224±0.08
	Recall	0.195±0.08	0.146±0.05
	F1	0.224±0.06	0.165±0.03
BLEU		0.041±0.02	0.038±0.01
Levenshtein Distance		1.450±0.30	1.420±0.20

Table 3. Comparison of drug effectiveness summary for DrugBank, ChatGPT, and our model.

Question	*Is Galantamine effective for improving cognitive function for Alzheimer's disease patients?*
DrugBank	Galantamine is a cholinesterase inhibitor used to manage mild to moderate dementia associated with Alzheimer's Disease. Galantamine is indicated for the treatment of mild to moderate dementia of the Alzheimer's type.
ChatGPT	According to clinical studies, Galantamine appears to have a moderate benefit in improving cognitive function in Alzheimer's disease patients, specifically on measures of attention, memory and thinking, as well as activities of daily living when compared to a placebo. However, it's important to note that results from the studies were mixed. …
Our method	The currently available evidence seems to support the use of Galantamine with a few contradictory evidence. We found five studies investigated the effectiveness of Galantamine: Two studies probably support the efficacy of Galantamine. The randomized control trial (PMID 15209643) found that after 5 months of treatment, Galantamine has a positive effect on ADL performance in patients with AD. The randomized control trial (PMID15525294) found that Galantamine is safe and effective for long-term treatment of mild-to-moderate AD. …

For the evaluation, we manually created drug questions instead of selecting questions and reference texts from the public medical QA dataset to avoid bias because ChatGPT has probably already been pre-trained on them. The results indicate that both our method and ChatGPT have advantages and limitations in summarizing evidence of drug effectiveness. ChatGPT can provide contextually relevant and personalized responses but struggles to extract detail or key information from specific clinical studies. On the other hand, our method can provide highly structured and explainable summaries of clinical studies. In light of these findings, it may be beneficial to use a combination of both summarization and neural language modeling methods to achieve a more comprehensive and accurate summary of the information. This would involve using the document summarization method to extract structured information and the ChatGPT model to generate a more human-like and nuanced summary. In practice, the choice between the two approaches will depend on the specific goals, resources of the research project, and the trade-offs between summary accuracy and comprehensibility. For medical evidence synthesis, combining summarization and neural language modeling methods can be particularly useful. Medical research involves complex terminology and concepts that require domain-specific knowledge to fully understand. Therefore, a summary that is both accurate and easy to understand is essential for medical professionals to make informed decisions.

4. Conclusions

Both ChatGPT and the proposed methods have advantages and limitations, with the former being able to mimic natural and human-like summaries and the latter being highly effective in extracting structured information with links to relevant studies. Nevertheless, combining both methods may lead to a more comprehensive and accurate summary of the drug effectiveness information. One limitation of our work is that we only sampled ten questions to evaluate the systems. A more comprehensive evaluation of large-scale datasets is needed. Next, we only evaluated the models on automatic metrics. Whether they are well-suited to evaluating zero-shot summaries is still being determined. While recent work has shown that classical reference-based scores, such as ROUGE, correlated with human preferences [16], we still need to conduct human evaluations to compare the outputs of models and collect human preferences for quality. We hope our study could encourage future work to address these limitations to further explore the potential of large language model learning on medication evidence synthesis.

Acknowledgments: This work was supported by the National Library of Medicine grant R01LM009886 and 4R00LM013001, National Center for Advancing Clinical and Translational Science award UL1TR001873, NSF CAREER Award No. 2145640, and Amazon Research Award.

References

[1] Afantenos S, Karkaletsis V, Stamatopoulos P. Summarization from medical documents: a survey. J Artificial intelligence in medicine. 2005;33:157-77.
[2] OpenAI. ChatGPT: Optimizing Language Models for Dialogue. Jan 9, 2023 ed. https://chat.openai.com/chat: OpenAI; 2023.
[3] Aydın Ö, Karaarslan E. OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. Emerging Computer Technologies 2. 2022:22-31.
[4] Demner-Fushman D, Lin J. Answering clinical questions with knowledge-based and statistical techniques. Computational Linguistics. 2007;33:63-103.
[5] Bellomo R, Bagshaw SM. Evidence-based medicine: classifying the evidence from clinical trials–the need to consider other dimensions. Critical Care. 2006;10:1-8.
[6] Pradeep R, Ma X, Nogueira R, Lin J. Scientific claim verification with VerT5erini. arXiv preprint arXiv:201011930. 2020.
[7] Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. Foundations Trends® in Information Retrieval. 2009;3:333-89.
[8] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:191010683. 2019.
[9] Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH). 2021;3:1-23.
[10] Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. 1995.
[11] Zhang J, Zhao Y, Saleh M, Liu P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. International Conference on Machine Learning: PMLR; 2020. p. 11328-39.
[12] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research. 2018;46:D1074-D82.
[13] Lin C-Y. Rouge: A package for automatic evaluation of summaries. Text summarization branches out 2004. p. 74-81.
[14] Papineni K, Roukos S, et al. Bleu: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting of the Association for Computational Linguistics 2002. p. 311-8.
[15] Yujian L, Bo L. A normalized Levenshtein distance metric. IEEE transactions on pattern analysis machine intelligence. 2007;29:1091-5.
[16] Deutsch D, Dror R, Roth D. Re-Examining System-Level Correlations of Automatic Summarization Evaluation Metrics. Seattle: Association for Computational Linguistics; 2022. p. 6038-52.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230348

Combining Sequence Similarity with Physicochemical Properties to Predict Binders for MHC-II Molecules

Ya-Lin CHEN[a,1]
[a]*Department of Biomedical Informatics and Medical Education, University of Washington, USA*
ORCiD ID: Ya-Lin Chen https://orcid.org/0000-0002-7046-827X

Abstract. Prediction of binders of the major histocompatibility complex class II (MHC-II) molecules is critical for T cell immunogenicity. As protein-protein interaction also relies on physicochemical properties, we aim to build a novel model combining sequence information and the physicochemical properties of proteins. Our research used data from the NetMHCIIpan 3.2 study. Features include BLOSUM50 and the physicochemical properties from iFeature Python package. We created a hybrid model of recurrent neural layers and feedforward layers. The final Area Under the Receiver Operating Characteristics (AUROC) on the test data was 0.755.

Keywords. MHC class II, MHC binding specificity, deep neural network

1. Introduction

The binding of Major histocompatibility complex class II (MHC-II) and peptides is a key step for T-cell recognition, triggering the defense against viral infection and other diseases. Identifying the peptides is crucial for understanding the pathophysiology of diseases and for designing therapeutics. Currently, several methods use sequence similarity data, such as BLOSUM matrix, to predict the binding affinity [1]. However, protein-protein interaction is also dependent on other factors such as physicochemical properties [2]. This research aims to combine sequential data and physicochemical properties to predict the binders of MHC-II molecules.

2. Methods

The dataset from NetMHCIIpan 3.2 consists of pairs of MHC-II molecules and peptides with their binding affinity [1]. We selected one partition of data and included only human data, MHC-II molecules with more than 20 peptides, with at least 4 binders with peptide length of 15, resulting in 87,052 pairs as train data and 21,535 pairs as test data. Binders were defined as those with < 500nM binding affinity [1]. The sequential features were generated using the BLOSUM50 matrix. We used iFeature Python library to generate

[1] Corresponding Author: Ya-Lin Chen, E-mail: end1859612@gmail.com.

the physicochemical properties [3, 4]. A hybrid model was created with a recurrent neural network (RNN) for sequential features whose latent output combined with physicochemical properties was to fed into a deep feedforward network. Neural network models were built with Tensorflow 2.10 and the code repository is available online.

3. Results

The percentages of binders in the train and test data were 43.42% and 40.34% respectively. The test performance is shown in Table 1.

Table 1. Test performance. Precision, Recall, and F1-score are macro averages. AUROC: Area Under the Receiver Operating Characteristics. RNN: recurrent neural network.

	BLOSUM50	**iFeature**	**BLOSUM50+iFeature**
Model	RNN layers	Feedforward layers	RNN + Feedforward layers
AUROC	0.740	0.711	0.755
Accuracy	0.752	0.713	0.750
Precision	0.742	0.705	0.746
Recall	0.740	0.711	0.755
F1-score	0.741	0.706	0.746

4. Discussion

Our design of the hybrid model takes into account both the sequential data and the physicochemical properties of proteins. However, one limitation of current research is that we only used one out of the five partitions of the data from NetMHCIIpan 3.2 [1]. We expect to improve the prediction as we include more data in the future [5].

5. Conclusions

This research combined BLOSUM50 and physicochemical properties to predict binders for MHC-II molecules. Using a hybrid deep neural network, the test AUROC reaches 0.755.

References

[1] Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018;154(3):394-406.
[2] Oscarsson S. Factors affecting protein interaction at sorbent interfaces. Journal of Chromatography B: Biomedical Sciences and Applications. 1997;699(1):117-31.
[3] Chen Z, Zhao P, Li F, Leier A, Marquez-Lago TT, Wang Y, et al. iFeature: a Python package and web server for features extraction and selection from protein and peptide sequences. Bioinformatics. 2018;34(14):2499-502.
[4] Chen D, Li Y. PredMHC: An Effective Predictor of Major Histocompatibility Complex Using Mixed Features. Front Genet. 2022;13:875112.
[5] Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics. 2013;65(10):711-24.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230349

Electromagnetic Fields Literature Analysis for Precision Medicine

Christopher HAWTHORNE [a*], Donal HARRISON [a*], Eva McEvoy [a*] and Guillermo LOPEZ CAMPOS [a,1]
[a] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast,UK*
ORCiD ID: Guillermo Lopez Campos https://orcid.org/0000-0003-3011-0940

Abstract. During the last century technological advances have increased the number of anthropogenic electromagnetic fields (EMFs) and therefore human exposures. In this work we have mined from more than 30,000 EMF-related publications the genes, diseases and molecular mechanisms associated with the exposure to six different subsets of EMFs. Results show 3653 unique disease MeSH terms and 9966 unique genes identified of which only 4340 genes are human. Overall, our approach highlights the molecular aspects of the increasing exposure to EMFs.

Keywords. Bioinformatics, Exposome, Electromagnetic fields, text mining, precision medicine

1. Introduction

The exposome [1] has become an increasingly relevant area of research during the last decade as the environmental counterpart of the genome to understand and explain the health and disease status of an individual. Among the different environmental sources of exposures, technological advances during the last century have caused an increase in the exposure to anthropogenic electromagnetic fields (EMFs). These are nowadays ubiquitous and are originated from multiple sources such as electrical wiring, appliances or telecommunications among others and have been object of an increasing regulatory and social scrutiny regarding their potential effects on human health. Recent controversy about the effects of mobile communications (and particularly 5G technologies) are a clear example of the relevance of research in this area and the increasing volume of publications. In this context, biomedical informatics and text mining offers the opportunity to analyze the corpus of literature associated with EMFs and extract information related with genes and diseases to investigate the molecular aspects related to EMF exposures. Here we present an in-silico analysis of EMF-related literature aiming to extract and compare molecular information and disease.

* Equal contributors
[1] Corresponding Author: Guillermo Lopez Campos, Wellcome-Wolfson Institute for Experimental Medicine, E-mail: g.lopezcampos@qub.ac.uk.

2. Methods

Bibliographic information for the six different EMF subcategories (direct current, electric currents, low frequency, mobile telecommunications, power frequencies and radio frequency) was extracted from the EMF-Portal (https://www.emf-portal.org/en) literature database in ".ris" format and processed to annotate selected documents with their PubMed identifiers (PMIDs). Then were incorporated to an MS Access database where they were further filtered to retain only articles in English. A second stage used PubTator [2] annotations to annotate the selected documents with genes and proteins, and diseases (MeSH terms). These annotations were further filtered to keep only human genes and proteins. These genes and proteins were then used in an additional analytical step, using gene set enrichment analyses to infer diseases and pathways (adj. p-value <0.05) associated with the different subcategories of EMFs, calculated using Enrichr [3].

3. Results and Discussion

A total of 32,986 English documents were retrieved, of which only 25,632 (77.7%) had a PMID and 11,667 (45.6% of PMID annotated documents) were uniquely associated to a single EMF subcategory. Results from PubTator annotations are presented in Table 1. In addition to disease inference, a total of 227 different KEGG pathways were associated as potential mechanisms affected by the EMFs.

Table 1. Results from text mining and gene set enrichment analyses

EMF category	No. of Documents	No. of Genes/proteins	No. of diseases (Found as MeSH)	No. Diseases (Inferred from DisGeNet)
Direct Current	6100	4683	2613	6075
Electric Current	6409	4747	2782	6103
Low Frequency	14732	7270	3317	6713
Mobile Comms	3301	2124	1362	5417
Power Frequency	4776	2824	1989	6200
Radio Frequency	7984	3917	2151	6713

Surprisingly, a large number of unique genes (9966) and diseases (3653) are found to be related with the different subcategories. However, of all the genes identified only 4340 have been annotated as human genes. There is an important redundancy in the genes and diseases found from the text and in the inferred diseases from the gene lists, this is not entirely unexpected as the corpus displays redundancy in the annotation of the different EMF subcategories. Overall, our approach has identified molecular elements related with the exposure to EMFs and these preliminary results require further biological in-depth analysis and will serve as the foundational element for a related graph knowledgebase.

References

[1] Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14: 1847–50

[2] CH Wei, A Allot, R Leaman, Z Lu PubTator central: automated concept annotation for biomedical full text articles. Nucleic Acids Res. 2019 Jul; 47(W1) W587-W593

[3] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 April; 14:128

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230350

A Reliable and Secure Method for Sharing Genomic Data

Dinis. B. CRUZ[a], João R. ALMEIDA[a,b,1], Jorge M. SILVA[a] and José L. OLIVEIRA[a]
[a] *DETI/IEETA, LASI, University of Aveiro, Portugal*
[b] *Department of Computation, University of A Coruña, Spain*

Abstract. Genomics has significantly impacted the field of medicine, with advances in DNA sequencing leading to personalized medicine and a deeper understanding of the genomic basis of various diseases. The ability to share genomic data is crucial for advancing this field and developing new approaches to understanding the genome. However, the sensitive nature of this data requires secure methods for protecting it during storage and transfer. In this paper, we present a new tool for the secure encryption and decryption of FASTA files without sharing a common secret and with a reduced number of shared keys between the pairs. Our proposal combines symmetric and asymmetric encryption techniques, including the AES (Advanced Encryption Standard) cypher and RSA (Rivest–Shamir–Adleman). The tool is fast, reliable, and secure, outperforming existing tools in terms of security and ease of use. This makes it a valuable solution for the secure sharing and use of sensitive genomic data, representing a significant advancement in the field of genomics.

Keywords. Genomics, DNA sequencing, data security, encryption, decryption, FASTA files.

1. Introduction

The technological advances in DNA sequencing led genomics to the frontlines of research in medicine, agriculture, and environmental sciences. They opened the door to the development of personalized medicine and the understanding of the impacts in climate change and pollution [1]. In this evolutionary path, sharing genomic data is crucial for advancing research and developing new approaches to better understand the genomes. Nonetheless, especially in the medical field, the nature of genomic data requires secure methods for protecting it during storage and transfer [2].

With this in mind, we propose SecureFASTA, a new tool for the secure encryption and decryption of FASTA files, which are commonly used to store DNA sequences and are frequently shared among researchers. This tool uses symmetric and asymmetric encryption techniques to ensure data confidentiality with a reduced number of necessary keys shared between the pairs.

2. Methods

The proposed tool aims to simplify the encryption process of genomic data represented in the Fasta format. It implements a strategy combining symmetric and asymmetric

[1] Corresponding Author: João Rafael Almeida, E-mail: joao.rafael.almeida@ua.pt.

encryption algorithms to optimize the procedure and reduce the number of necessary keys shared between the pairs.

The proposed tool uses AES cypher to encrypt the Fasta files. AES-256 is a widely used and efficient symmetric cypher that is hard to crack using brute force and dictionary attacks when using random keys. The tool generates a random secret key for each file. The secret key is then used to initialize an instance of the AES cypher. The input file is read, encrypted, and a new file is written. The generated key is then encrypted using asymmetric encryption. We used RSA (Rivest–Shamir–Adleman) cypher [3] since it is widely used for secure data transmission.

3. Results and Discussion

We validated the implementation of the proposed tool using synthetic Fasta files containing nucleotide and protein sequences of various sizes. In terms of security, the encrypted files produced by this tool were resistant to tampering. Any attempt to modify the encrypted file resulted in a decryption error, indicating that the tampering had been detected.

Regarding security, our method follows best practices outlined in the GA4GH File Encryption Standard [4]. This standard recommends using authenticated encryption to ensure the confidentiality, integrity, and authenticity of the data and using cryptographic key management systems to store and manage encryption keys securely.

The file's checksum generated by SecureFASTA allows the receiver to confirm its integrity before decrypting it. The algorithm used for this was SHA-256 since it is resistant to collisions. In this context, a collision is when two distinct pieces of data in a hash table share the same hash value.

In this work, we proposed a valuable tool for ciphering Fasta files. Its features include the use of both symmetric and asymmetric encryption and a checksum function that allows the recipient to verify the integrity of each file before deciphering.

Acknowledgments

This work has received funding from the EC under grant agreement 101081813, Genomic Data Infrastructure.

JMS and JRA have been funded by FCT (Foundation for Science and Technology) under the grants SFRH/BD/141851/2018 and SFRH/BD/147837/2019, respectively.

References

[1] Silva JM, Almeida JR. The value of compression for taxonomic identification. In: 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2022. p. 276-81.
[2] Bradley T, Ding X, Tsudik G. Genomic Security (Lest We Forget). IEEE Security & Privacy. 2017;15(5):38-46.
[3] Rivest RL, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM. 1978;21(2):120-6.
[4] Rehm HL, Page AJ, Smith L, Adams JB, Alterovitz G, Babb LJ, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell genomics. 2021;1(2):100029.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230351

Mapping Exposome Derived Phenotypes into SNOMED Codes

Christopher HAWTHORNE [a*], Luis MARCO RUIZ [b*] and Guillermo LOPEZ CAMPOS [a,1]
[a] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast,UK*
[b] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
ORCiD ID: Guillermo Lopez Campos https://orcid.org/0000-0003-3011-0940

Abstract. Human phenotypes define the healthy or diseased status of an individual and they arise from the complex interactions between environmental and genetic factors. The whole set of human exposures constitute the human exposome. These exposures have multiple sources including physical and socioeconomic factors. In this manuscript we have used text mining techniques to retrieve 1295 and 1903 Human Phenotype Ontology terms associated with these exposome factors and we have subsequently mapped 83% and 90% of the HPO terms respectively) into SNOMED as a clinically actionable code. We have developed a proof-of-concept approach to facilitate the integration of exposomic and clinical data

Keywords. Bioinformatics, Exposome, precision medicine, SNOMED-CT, Human Phenotype Ontology.

1. Introduction

Human phenotypes define the healthy or diseased status of an individual and they arise from the complex interactions between environmental and genetic factors. In the last couple of decades, great efforts have been made to improve the tailoring of medical practice to individuals, mostly through an improved understanding of the genome of an individual. More recently, with the development of precision medicine approaches, the exposome [1], defined as the whole set of life-through exposures of an individual, has also started to be considered. This "whole set" definition makes the exposome a broad and complex element that covers exposures to biological, chemical, physical and psychosocial agents. The contributions to exposomic research from each of these components is variable and so far, the focus has been mostly on understanding the chemical and biological domains, whereas an increasing interest is also noticeable in psychosocial aspects and the social determinants of health. Biomedical informatics play a key role in the development of efficient solutions that facilitate the analysis and integration of these data in multiple scenarios. In this manuscript we have combined computational exposomic research on physical and socioeconomic factors related

* Equal contributors

[1] Corresponding Author: Guillermo Lopez Campos, Wellcome-Wolfson Institute for Experimental Medicine, E-mail: g.lopezcampos@qub.ac.uk.

literature with SNOMED annotations as an approach for the future potential validation of the results and the integration of clinical data.

2. Methods

Bibliographic contents from the title and abstract of PubMed indexed literature around environmental 16 different physical factors (PF) ("lighting", "noise", "electromagnetic fields"…) and 7 different socioeconomic factors (SEF) (including "income", "poverty" and "educational status" among other) were annotated using ONASSIS tool to identify human phenotypes as previously described [2]. Phenotypes were then mapped to SNOMED-CT using the UMLS Terminology Service checking the available mappings in SNOMED-CT for each of the HPO codes representing the human phenotypes extracted from PubMed. The US 2022 SNOMED-CT version was used.

3. Results

A total of 5844 (4541 for PF + 1303 for SEF) different HPO terms were retrieved from the initial analyses. After filtering out phenotypes appearing in multiple factors 1294 (28%) and 909 (70%) unique HPO terms for physical and socioeconomic factors respectively remained in the dataset.

SNOMED mapping resulted in 1071 (83%) HPO terms mapped to different SNOMED codes for the physical factors dataset and 822 (90%) HPO terms mapped for the socioeconomic factors dataset. Interestingly, although initially a smaller set, SEF related HPO terms led to a more diverse set of SNOMED codes. A total of 8439 SNOMED codes (4523 for PF + 3916 for SEF) were mapped. The majority of the HPO terms were associated with more than one SNOMED code (min=0, average=3.4, max= 23 for the PF and min=0, average=4.3, max= 23 for the SF).

4. Conclusion

Socioeconomic related phenotypes are more easily mapped into a more diverse set of SNOMED-CT codes than those derived from those associated with the physical factors of the exposome. We have been able to develop a proof-of-concept approach to integrate exposomic related knowledge extracted from the literature with clinical vocabularies that might facilitate the validation of the research hypothesis derived from the literature using real world clinical data.

References

[1] Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14: 1847–50

[2] Hawthorne C, Quigley N, McClements C, Lopez-Campos GH. Construction of a Physical Factor Resource for Exposome Informatics Research. Stud Health Technol Inform. 2021 May; 281:1079-1080.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230352

The RCX Extension Hub: A Resource for Implementations Extending the R Adaption of the Cytoscape Exchange Format

Florian AUER[a,1], Nadja BRAMKAMP (BECKMANN)[a], Simone MAYER[a], Dominik MÜLLER[a] and Frank KRAMER[a]
[a] *IT-Infrastructure for Translational Medical Research, University of Augsburg, Germany*
ORCiD ID: Florian Auer https://orcid.org/0000-0002-5320-8900, Nadja Bramkamp (Beckmann) https://orcid.org/0000-0001-8436-5274, Simone Mayer https://orcid.org/0000-0002-7825-5738, Dominik Müller https://orcid.org/0000-0003-0838-9885, Frank Kramer https://orcid.org/0000-0002-2857-7122

Abstract. Public repositories provide access to biological networks for investigations, and subsequently serve to distribute the network encoded biomedical and even clinically relevant results. However, inclusion of complementary information requires data structures and implementations customized to the integrated data for network representation, usage in supporting application, and extending analysis functionality. Partitioning of this information into individual aspects of a network facilitates compatibility and reusability of the network-based results, but also requires support and accessibility of the extensions and their implementations. The RCX extension hub offers overview and access to extensions of the Cytoscape exchange format implemented in R. The hub supports the realization of self-customized extension through guides, example implementations, and a template for the creation of R extension packages.

Keywords. Bioinformatics, biological networks, R, data models, extensibility.

1. Introduction

Biological networks capture complex association between biological entities, mechanism associated with their function and their involvement in biological processes. Within analyses the networks are enriched with diverse information, including their visualization, provenance history, and usage in additional applications. The NDEx platform [1] is an online commons for biological networks that uses the Cytoscape exchange (CX) format for transmission: It separates a network into its different aspects to modularize its contents. Within the statistical programming language R the *ndexr* package [2] allows querying the NDEx platform, retrieve and upload networks, manage their properties, and share them with specific people and groups. Thereby, the *RCX* package [3] adapts the CX data structure to standard R data types and implements methods for conversion, validation, summarization, visualization, and extension.

[1] Corresponding Author: Florian Auer, IT Infrastructure for Translational Medical Research, Alter Postweg 101, 86159 Augsburg, Germany; E-mail: florian.auer@informatik.uni-augsburg.de.

2. Methods

The RCX extension hub organizes the information around the development of custom aspects and their implementation. The hub therefore is structured in three sections: an overview over existing extensions, guiding the adaptation process through tutorials, and supporting the implementation with a package creation template. The RCX extension template can be used to create all required files, generic functions, and setup to build a functional package. Newly created extension packages can be registered raising GitHub issues through the hub.

3. Results

The RCX extension hub supports developers with their implementation by providing detailed guides and highlighting the requirements. Furthermore, the officially provided extensions serve as exemplary implementations illustrating representative use cases.

4. Discussion

In contrary to object-oriented programming languages, R approaches data in a vector and table-oriented manner and therefore requires an adaptation of the JSON encoded data objects contained in the CX networks. Nevertheless, this additional work required for the data transformation is paid off by the afforded accessibility to standard bioinformatics analysis tools and workflows provided by the R language. With this approach implementation of the different versions of the same aspect can also be easily managed.

5. Conclusion

The RCX extension hub forms a resource for the management of R implementations of custom CX aspects. It provides available extensions together with their intended applications, along with guides, examples, and templates for the development of own extension packages, and makes the hub a valuable resource and contact point for R-based network biology. All necessary source code of the official extensions is publicly available through the provided links and the RCX extension hub itself is hosted at https://frankkramer-lab.github.io/RCX-Extension-Hub.

References

[1] Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, Ono K, Miello C, Hicks L, Szalma S, Stojmirovic A. NDEx, the network data exchange. Cell systems. 2015 Oct 28;1(4):302-5.. doi:10.1016/j.cels.2015.10.001.

[2] Auer F, Hammoud Z, Ishkin A, Pratt D, Ideker T, Kramer F. ndexr—an R package to interface with the network data exchange. Bioinformatics. 2018 Feb 15;34(4):716-7. doi:10.1093/bioinformatics/btx683.

[3] Auer F, Kramer F. RCX—an R package adapting the Cytoscape Exchange format for biological networks. Bioinformatics Advances. 2022;2(1):vbac020. doi:10.1093/bioadv/vbac020.

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

Subject Index

3D/2D image analysis 895
5G 1002
accelerometry 1025
acknowledgement of CDSS alerts 511
activity trackers 876
ADDIE model 222
adherence 23
adolescent mental health care 1021
adolescents 242
adoption 227
adoption behavior 123
adverse drug reaction(s) 382, 384, 396, 551
adverse events 157
aerobic exercise 1023
AFO 1029
age-related macular degeneration 947
agent-based modelling 297
agile development 753
AI applications 676
AI-based applications 678
air pollution exposure 901
AKTIN 362
alarm fatigue 197
alarm safety 197
alert 197
alopecia 396
Alzheimer's disease 866, 937
ambient assisted living 621
animal owners 1017
ankle-foot orthosis 962, 1029
annotation 586, 803
anonymization 28
antibiotics 905
app development 366
app store 370
archiving 721
Armenia 476
artificial intelligence (AI) 103, 143, 177, 267, 346, 348, 408, 428, 488, 536, 571, 891, 909, 927, 972, 1007
ASD 716
assistive technology 504
asthma in children 913
ATC 711
atrial fibrillation 977
AUCMEDI 917
audio-visual evoked potential 257
augmented reality 413
austim spectrum disorder 716
autism spectrum disorder 123, 403, 1047
automated detection 232
automated diagnosis 615
automation 586
BabelDr 813
ballistocardiography 1031
Bayesian networks 987
benchmarking 167, 706
BERT 403, 783, 831
bias 488
big data 428, 827, 866
bigdata platform 372
BioBank System 392
bioethical issues 656
bioinformatics 1042, 1069, 1073, 1075
biological networks 1075
biological ontologies 745
biomedical data 147
biomedical ontologies 731
biometric authentication 257
biometry 438
biosamples 93
blood vessel segmentation 581
bone mineral density 909, 911
boolean minimization 302
BPMN 68
Braden scale 350
brain tumor 1052
breast cancer 396
breastfeeding 360
bwHealthApp 1019
Canada 783
cardiac rehabilitation 8
cardiology 819

cardiovascular disease(s) 63, 603
CARE 53
care continuity 886
case report forms 137
catalogs 438
CDSS and diabetes management 511
cervical intraepithelial neoplasia (CIN) 615
challenges 125
chatbot 157, 376, 478, 957
ChatGPT 1062
chronic disease 1019
chronic lymphocytic leukemia 384
chronic obstructive pulmonary disease 982
citizen expectation 468
claims data 3
classification 151, 917, 927
client information system 504
clinical 571
clinical adoption meta-model 227
clinical data reuse 247
clinical data warehouse 202, 312, 342
clinical decision support 516, 531, 596, 607, 1013
clinical decision support system(s) 197, 360, 511, 521, 551, 591, 611, 656
clinical guideline 605
clinical informatics 476
clinical notes 803
clinical physiology 759
clinical prediction model 129
Clinical Quality Language (CQL) 541
clinical research forms 247
clinical research infrastructure 93
clinical study 745
clinical trials 177, 721
clinical-administrative documentation 262
clustering 1052
CNN 364
co-creation 433
coaching 332
cognitive OR 149
cohort study 93, 716, 871
collaboration 671
colorectal cancer screening 394
common data elements 277
common data model(s) 322, 691
communication 139, 418, 636, 813
communities of practice 472
community-based participatory research 500
comorbidity 38
competencies 438, 472
complexity 506, 546
computable knowledge 541
computed tomography 63, 909
computer and information sciences 58
computer-assisted decision making 611
Computerized Clinical Decision Support Systems (CCDSS) 192
connected device 23
connected health 8
constrained optimization 773
container 58
context-aware operating room 149
continual training 763
continuing professional development (CPD) 496
continuity of patient care 141
continuous monitoring 1013
contrast 952
controlled natural language 829
controlled terminologies, ontologies, and vocabularies 706
controlled vocabulary 696
conversational agent 157, 376
convolutional neural network(s) (CNN) 526, 576, 615, 927, 947
core data set 749
corpora 743
corpus 586
cost-effectiveness 755
COVID 490
COVID-19 93, 463, 482, 496, 521, 536, 546, 743, 747, 783, 861, 891, 893, 901, 907, 917, 1015
COVID-19 symptoms 833
CQL artefacts 541
critical pathway 636
CSV 43
CT 1027
CT angiography 952
curricula 438
cyber security 337

CycleGAN 1027
cystic fibrosis 480
data analysis 282, 626, 876, 1042
data augmentation 927
data bias 428
data capture 829
data collection 696
data enrichment 147
data exploration 749
data governance 53, 312
data harmonization 751
data integration 33, 747, 749
data life cycle 757
data modeling 312
data models 1075
data protection 48, 317
data qualification 153
data quality 38, 167, 202, 217, 262, 272, 292, 302, 322, 380, 382, 871
data quality management system 322
data recording 1031
data repositories 721
data reuse 93, 147, 502, 721, 851, 856
data security 1071
data sharing 53, 58, 68, 252, 362, 386, 388, 721
data standardization 153
data transformation 43
Data Vault (DV) 753
data visualisation 354
data visualization 474, 626
Data Warehouse (DWH) 753
data work 262
datasets 133
datasets as topic 696
DBLP 162
de-identification 28
decentralized monitoring 252
decision making 605
decision support 177, 601
decision support systems 571, 726
deconvolution 73
decryption 1071
deep embedding 237
deep learning 63, 182, 364, 526, 576, 581, 609, 613, 615, 819, 911, 917, 932, 977, 997, 1033
deep neural network 1067
deep phenotyping 1037
Delphi study 346
dementia 113, 646
demonstration 494
dentistry 905
depression 817
design 98, 881
diabetes 521, 1009
diabetes mellitus 636, 856
diagnosis classification 815
diagnosis time 103
dialog data analysis 651
dietitian 398
digital 453
digital competence 494
digital divide 428
digital ecosystem 23
digital game 443
digital health 408, 967
digital health application 423
digital health integration 1019
digital health intervention 641, 1009
digital health technology 222
digital prevention 937
digital skills 504
digital surveillance systems 861
digital twin 1007, 1052
digital work-tool 187
digitalization 496
disease database 701
disease prediction 609
distributed business process 252
distributed systems 362
DNA sequencing 1071
docetaxel 396
documentation burden 661
documentation of nursing 531
documentation quality 398
domain knowledge 352
drug drug interaction 726
drug safety 551
drug-drug interactions 384
DXA 911
e-coaching 922
e-learning 443, 496
e-prescription 207
eating disorders 1033
ECG 33, 182
ecosystems 886
eCRF 380

edge computing 1002
education 448, 498
effective engagement 1009
eHealth 282, 332, 354, 490, 686, 778, 922, 1009, 1021
EHR data 202
EHR transition 448
elderly 222
elderly care 504
electrocardiogram 977
electrocardiography 1025
electrodermal activity 73
electroencephalogram 232
Electroencephalography (EEG) 257, 992
electromagnetic fields 1069
electronic health data 48
Electronic Health Records (EHRs) 141, 192, 242, 247, 262, 272, 344, 348, 352, 362, 364, 378, 388, 448, 490, 561, 603, 609, 613, 788, 827, 837, 899, 1021
electronic patient record systems 531
electronic prescribing 13
electronic prescription 899
electrophysiology 368
emergency 813
emergency disaster management 881
emergency medicine 815
emergency preparedness 907
emergency room decision support 905
emerging infectious diseases 861
emotion detection 73
emotion recognition 992
encryption 1071
endometriosis 103
endoscopy 576
engagement 967
ensemble learning 917
entity classification 808
environmental data 913
epidemic and pandemic response 83
epidemiology 438
epidemiology informatics 866
epilepsy 232, 1025
ePrescription 551
ergonomics 636
errors 242
estimation 182
ethical aspects 676
ethics 488
ETL 292, 751
European health data space 48, 131
evaluating explanations 1057
evaluation 378
event detection 798
evidence 423
evidence-based guidelines 531
evidence-based procedures 531
exertion level 1023
experiencer 1037
experiences 641
expert interviews 666
explainable AI 1057
Explainable Artificial Intelligence (XAI) 571
exploration 747
exposome 1069, 1073
extensibility 1075
external validation 33, 129
eye-tracking 631
facial genetics 932
FAIR 53, 368, 386, 390, 721, 741, 871
FAIR data 108
FAIR principles 63, 93
fairness 488
fast health interoperability resources 43
FASTA files 1071
feature importance 903
feature selection 566
federated data capture 137
federated data network 307
federated learning 516
female reproductive health 103
few-shot learning 932
FHIR 43, 68, 133, 135, 317, 358, 372, 386, 541, 937
Finland 227
five rights model 531
foot drop 962, 1029
form 745
framework 68, 118
fusion 615
Fuzzy Cognitive Map (FCM) 992
fuzzy logic 992
gait analysis 962
gamification 332, 443

GATE 793
GDPR 143, 267, 488
gene expression data 1047
general practitioner 856
generative adversarial networks 927
genetic algorithms 566
genome 386
genomics 1071
German medical informatics initiative 835
Germany 125
glaucoma 581
glioblastoma 972
google play 370
google trends 861
gradient boosted trees 952
graph database 749
graph neural networks 561, 603
Guided LDA 783
gut-brain axis 1007
hand 911
health 666
health and social care 494
health apps 423
health care professionals 141
health communication 463
health data 131, 302, 913
health data access body 131
health data sharing 28
health emergencies 408
health equity 428
health informatics 671
health information interoperability 899
health information platform 83
health information retrieval 833
health intervention 463
health issue 506
health issue network 506
health IT 907
health monitoring 561, 1002
health services research 362
health technology assessment 656
health-content 743
healthcare 38, 496, 987
healthcare analytics 153
healthcare data 541
healthcare leaders 678
healthcare professionals 151, 676
healthcare standards 292, 372
healthcare workforce 472
heart failure 755
heart sound 526
height 364
helpline 474
high-impact health events preparedness 83
histology image 615
HIV 500
HL7 372
HL7 CDA 899
HL7 FHIR 113, 272, 292, 706, 749
home care 453
home-based 682
hospital 344
hospital data 282
hospital information systems 141
hospital management 486
hospital readmission prediction 556
hospitals 617
human activity 987
human digital twin 841
human factors 192
human phenotype ontology 1073
human-centered AI 556
human-centered design 646
human-computer interaction 631
human-robot interaction 651
ICD-11 731
ICF 763
ICU 237, 366
ICU—intensive care unit 536
image analysis 932
image segmentation 947
images 93
imbalanced data 932
immersive learning 433
implementation 227
implementation science 596
implementation strategy 736
incident report 374
inertial measurement unit 962
infection-lung ratio 917
infodemic 891, 893
informal caregiving 922
informatician 671
informatics 207
information dissemination 141
information driven care 346

information extraction 831
information model 829
information needs 881
information networks 362
information technology 490
information-driven care 346
infrastructure 177, 841
injury prediction 118
innovation in digital health 736
inpatient cancer mortality prediction 821
integrated health care systems 886
integration 125
intelligent eLearning 482
intelligent phonocardiography 526
intelligent workflow 661
interdisciplinary healthcare 556
interface design 380
international exchange 498
internet of medical things 937
internet of things 337
internet search engines 861
interoperability 3, 8, 43, 88, 108, 113, 133, 135, 139, 145, 317, 366, 388, 541, 671, 706, 711, 751, 846
interpretability 992
interpreting 813
interview 287
intraclass correlation coefficient 962
introduction 1011
IoMT 937
IoT 337, 666
IPHIE 498
ISO 14721 721
IT systems 332
IT-architect 671
Italy 297
k-anonymity 28
K-means 1052
Kanta Services 227
key performance indicators 282
knee 909
knowledge 141
knowledge commons 480
knowledge graphs 147, 701, 747
knowledge representation (computer) 611
knowledge reproduction 743
knowledge support 531
laboratory test 88
language models 743, 763, 819
large language model 833
large scale data sharing 307
learning analytics 463
learning health system 145
learning networks 480
learning objectives 438
lecture 484
literature database 162
LMICs 476
logistic regression 952
LOINC 88
long short-term memory (LSTM) 861
longitudinal data 366
machine learning (ML) 33, 38, 73, 123, 127, 182, 232, 237, 350, 516, 536, 571, 596, 691, 726, 788, 817, 897, 901, 903, 909, 972, 1023
machine learning methods 1027
machine-learning 83
Markov chain 755
Markov model 987
masked language model 609
maturity model 907
me-to-we design 621
mechanical ventilation 566
mechanical ventilators 680
MeDaX 147
medical appointment management 376
medical appointment scheduling 376
medical communication 823
medical concept mapping 825
medical data 125, 755
medical data integration 691
medical data management 376
medical data models 137
medical device(s) 337, 413, 626
medical education 443
medical eponymy 808
medical imaging 997
medical informatics 438, 498, 741
medical questionnaire 376
medical record 13, 398
medical secretaries 262
medical semiology 482
medical texts 831
medication 207, 287, 711
medication management 327

medicine 438, 731
medico-administrative database 394
mediterranean buffalo 895
menstrual health 103
mental health 356
mentoring 484
meta-learning 932
metadata 302, 368, 751
metadata annotation 108
MHC binding specificity 1067
MHC class II 1067
mHealth 370, 468, 500, 646, 922, 1017, 1019
mHealth prescription 468
mHealth: remote monitoring 1013
microneurography 368
MIMIC 98, 566
MIMIC IV 237
MIMIC-III 561
Mirth Connect 43
mobile 197
mobile app
mobile application(s) 222, 492, 778, 957
mobile app(s) 172, 370
mobile health 370, 957
model-driven intervention planning 937
modeling 726
modelling 881
modular architecture 212
monitoring 342, 1025
mortality prediction 561
mRNA vaccines 798
mucoviscidosis 358
multidimensional model 546
multidisciplinary team 605
multimorbidity 113
N-of-1 480
named entity normalization 837
named entity recognition 831
national health insurance reimbursement 78
national survey 242
natural language processing 743, 768, 788, 798, 803, 815, 817, 827, 831, 835, 1062
NCBI 1047
negated phenotype 1037
Neo4j 701
network university medicine 741
neurodevelopmental disorder 716
neuroradiomics 972
neurosurgery 972
NLP 392, 773, 793
noise 977
non-English EMR clinical text processing 821
nursing 344
nursing care 470
nursing documentation 778
nursing home(s) 350, 502
nursing informatics 438, 591
nursing innovation 684
nursing quality dashboard 617
nursing robotic 470
nursing sensitive indicators 617
nursing technology 684
nutrition care process 398
nutrition measurement 1033
OAIS 721
obesity 478
observables 759
observational study 247
obstetrics and gynecology (specialty) 571
occupation 151
OHDSI 3, 711
older persons 453
OMOP 358, 711
OMOP CDM 3, 751, 755, 899
oncology 18, 125, 277, 605, 817
online learning 408
ontology 759, 793, 905
open data 851
open metadata Markup Language (odML) 368
open source 43, 68, 147
ophthalmology 581
opioid use disorder 897
orpea 502
osteoporosis 909
outlier detection 88
outpatient 1025
pain 368
palliative care 686
pandemic preparedness 891, 893
pandemic response 891, 893

parallel convolutional neural network 526
Parkinson's disease 127
participation 470
participation management 684
participatory design 641
participatory research 23
patient care planning 886
patient data management system 358
patient empowerment 418, 458
patient generated data 135
patient generated health data 372
patient journey 8
patient mobility 297
patient modeling 237
patient monitoring 172
patient portal(s) 490, 1021
patient reported outcome 135
patient safety 374, 413, 680
patient similarity 1037
patient telemonitoring 332
patient trajectories 609
patient-accessible electronic health record (PAEHR) 356
patient-accessible electronic health records 242
patient-contributed data 8
patient-controlled real world data 480
pediatric obesity 957
pediatrics 516, 651
pen injector 23
person-centered care 187
personal health record 145
personalized digital health 212
personalized nudge 876
persuasive design 922
PET 972
Petri Nets 506
phantom limb pain 682
pharmacovigilance 382, 601
pharmacy 360
pHealth 332
phenotyping 768
Phlebitis 374
physical activity 478, 876
physician satisfaction 686
physicians 496
PICS 366
pictographs 823
platform 98
polyps 576
polyvalent immunoglobulins 342
population health management 886
Portugal 48
postoperative infections 348
postural balance 1029
powered by communities 83
practice guidelines as topic 611
precision 591
precision care 937
precision livestock farming 895
precision medicine 1069, 1073
prediction 348, 967
prediction model 98, 139, 903
prediction modelling 1057
predictions models 103
pregnancy 360
premature birth 571
pressure ulcers 350
prevalence study 913
primary care 856, 942
primary healthcare 1011
primary use of data 48
privacy 212, 656
privacy preservation 143, 267
privacy preserving technique 28
privacy-preserving protocols 516
process assessment 344
process mining 18
profiling 113
prognostic factors 901
prompting 819
protocol 913
prototyping 626
pseudonymization 317
psychiatry 474, 793, 851
public attitudes 783
public health 463, 866, 903
public health informatics 881, 895
PubMed 162
pulmonary rehabilitation 458
qualitative analysis 982
qualitative research 478
qualitative study 676, 678
quality indicator(s) 167, 696
quality management 871
quality of life 135
R 1075

radiograph 911
radiology 277
radiomics 972
random forest 952, 1047
rare disease(s) 292, 358, 607
rare genetic disorders 932
RDA FAIR data maturity model 63
RDF 133
real world data 88, 307
real-time data 651
real-world data 18
recall 591
recognition 987
recombinant human growth hormone 23
recommendations 332
Reddit 783
registration practices 262
registry 167
regulation 488
rehabilitation 763
reimbursement 423
representation learning 352, 609
reproducibility 58
research 851
research data management 721
residency education 496
resilience 907
resources utilization 613
respiratory rate 1025
respiratory system 696
results communication 277
retrieving 98
retrospective study 342
rheumatoid arthritis 768
risk factor 521
risk management 157
risk of bias 586
routinely collected health data 362
rule-based expert system 332
RWD 711
RxNorm 711
safety 157
sample size 182
SARS-CoV-2 536
satisfaction 813
scalability 841
scientific societies 438
score-based generative model 1027
screening 909
secondary use 272, 788
security 212
segmentation 581
seizure prediction 232
seizures 1025
selection bias 127
self-administered 682
self-efficacy 661
semantic annotation 793
semantic enrichment 757
semantic health data 829
semantic interoperability 78
semantic metadata 108
sensitive data 143, 267
sentiment analysis 651
service realisation 736
severity 917
sex and gender 757
Shapley values 1057
shared information 287
shared information lists 327
shared medication list 207
signal assessment 601
Sildenafil 866
simple steps 390
simulation 202
simulation process 297
small cell lung cancer 18
smart cameras 895
smart hospital 661
smart sensors 1007
smart watch 876
smart wearables 1002
smartphone apps 468
SNOMED CT 731, 825, 829, 1073
social activity 453
social determinants of health 217
social isolation 453
social listening 891, 893
social media 403, 478, 502, 783, 861, 893
social media monitoring 891
socio-technical system 656
sociotechnical perspective 287
software 1031
Spanish 500

spatial accessibility 297
spectral domain optical coherence tomography 947
speech recognition 778
SPHN 133
SQL 98
stacking ensemble 129
standard(s) 133, 846
standing orders 13
statistical test 1047
stroke 172
structural topic modeling 798
structured reporting 277
study monitor 358
subphenotyping 897
substance use 967
suicide 474
summarization 1062
summer schools 484
supervised learning 773
surgery 418
surgical data science 149
Swedish internet forum 798
Swedish tweets 798
symposia 484
syndrome 607
synthetic data 352
synthetic data 53, 378
synthetic images 927
systematic review(s) 162, 591, 586, 680
systematized nomenclature of medicine clinical terms 78
systemized nomenclature of medicine clinical terms 372
T2DM 492
tablet 453
tasks 492
taxonomy 893
TEDIS 716
telehealth 1017, 1021
telemedicine 656, 942, 1015
teleradiology 997
telerehabilitation 982
template 58
Temporal Association Rules (TAR) 546
terminology 759
text classification 803
text corpus 835
text mining 1069
textile electrodes 682
time and frequency features 257
time-domain features 73
time-sensitive data imputation 237
token economy 486
tool 494
topic modeling 502
topic modelling 403, 783
touch screen 631
traditional 731
traffic accident simulation 118
training 448, 476
transcriptomics 1042
transfer learning 396
transformer models 808
translational health informatics 386
treatment decision 18
treatment platform 937
treatment trajectories 755
trust 83
trustworthy information 83
Twitter 403, 502, 833, 861
type 2 diabetes 38
type 2 diabetes mellitus 492
U-Net 581
umls 823
unsupervised learning 833
usability 192, 492, 500, 631, 646, 686
usability assessment 684
usability engineering 626
usability-test 680
use 227
usefulness 678
user involvement 641
user participation 661
user requirements 666
user satisfaction 192
user-centered design 641, 680, 778
user-centred design 327, 354
vaccine hesitancy 463, 798
vaccine mandate 783
validation 68
validation protocol 380
variable importance 1057
vascular diseases 1013
veterinary medicine 1017
video 631
video consultations 942, 1011

village health volunteer 222
virtual reality 413, 458
virtual/augmented reality 433
visual analytics 474, 1042
visualization 252
vital signs 145
waiting list data 846
weaning 566
wearable devices 1013, 1023
wearable electronic device 1025
web applications 871
WebAssembly 997
weight 364
welfare technology 621
well-being 666
what matters to patients 187
workflow 881
workshop 388
wound imaging 927
wounds and injuries 354
X-ray retakes 486
XAI 613
XGBoost 257
young scientists 484
zero-shot learning 833

This page intentionally left blank

Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

Author Index

't Hart, H.H. 815
Aakenes Vik, K.T. 488
Abdulnazar, A. 825, 837
Abi Khalil, C. 591
Abovarda, A. 496
Absetz, P. 1009
Accordino, F. 297
Adedeji, T. 388
Agastinose Ronickom, J.F. 73, 232, 257, 1047
Ågerfalk, P. 922
Åhlfeldt, R.-M. 356
Ahlin, K. 494, 666
Ahmad, A. 922
Ai, M. 486
Aittola, K. 1009
Akpinar, A.E. 113
Aktin Research Group 362
Al Suwaidi, H. 463
Albayrak, E. 470
Albrecht, U.-V. 370, 423, 1031
Alhaskir, M. 1025
Alkarkoukly, S. 43
Almeida, B. 337
Almeida, F. 516
Almeida, J.R. 1071
Al-Nawaiseh, S. 947
Alves, P. 354
Amariei, A. 851
Amaya-Rodriguez, I. 521
Ameen, S. 428
Ameijeiras-Rodriguez, C. 516
Amirahmadi, A. 609
Amitrano, F. 962, 1029
Ammon, D. 317
Anders, C. 498
Andersen, A. 876
André, E. 917, 932
Andrikopoulou, E. 541
Anton, G. 93
Antoniou, P. 433
Anywar, M. 541
Aouina, O. 793
Arbous, S. 348
Armellin, M.T. 763
Armengol De La Hoz, M.A. 521
Årsand, E. 478, 841
Arvanitis, T.N. 337
Arzehgar, A. 987
Athanasiou, A. 433
Auer, F. 1075
Ayumba, E.M. 482
Azé, J. 561, 773
Azqueta-Alzúaz, A. 153
Babic, A. 526
Bae, Y.S. 372, 1015
Baez, P. 500
Balestra, G. 566
Balke, D. 871
Bamidis, P. 433, 1052
Bamidis, P.D. 376, 571
Banaye Yazdipour, A. 443
Baran, J. 636
Bärkås, A. 356
Barkhau, C. 927
Barkhausen, J. 952
Barrett, J.W. 601
Bartschke, A. 741
Basri, M.A. 861
Bassetti, M. 380
Bathelt, F. 3, 711, 751, 753
Bauer, B. 917
Bauer, J. 1025
Bauer, M.P. 815
Baumberger, D. 151
Baumgartner, M. 803
Becker, L. 162
Beelich, H. 757
Begerau, H. 317
Belmin, J. 350
Ben Said, M. 716
Benamouzig, R. 394
Bender, T. 977
Benis, A. v
Berentzen, M. 162
Bergelt, C. 757

Berger, U. 438
Bergier, H. 768
Bergmo, T.S. 207
Bernard, G.R. 833
Berntsen, G.R. 187
Bertelsen, P. 262
Beume, D.P. 470, 684
Beuscart, J.-B. 502, 636, 856
Beutter, C.N.L. 135
Bevolo, V. 957
Bey, R. 202
Beyan, O. 43, 125, 1027
Bhattacharyya, S. 217
Bialke, M. 68
Biancardi, A. 962, 1029
Bickenbach, J. 366
Bickmann, L. 33, 182
Bienzeisler, J. 362
Billis, A. 376, 571
Biloborodova, T. 615
Bischof, A. 952
Bittencourt, I. 172
Björk, J. 609
Björkquist, A. 682
Blanchard, A. 337
Blasini, R. 358
Blayney, D. 817
Blease, C. 242
Blondon, K. 197
Blumenstock, M. 137
Blusi, M. v, 453
Boaglio, A. 386
Boeker, M. 272, 835
Bonacina, S. v, 731
Bonan, C. 743
Bonetto, M. 846
Bonotis, P. 384, 686
Bories, M. 342
Borlinghaus, H. 1019
Borycki, E. 907
Borycki, E.M. 881
Bosselmann, C. 611
Bott, O.J. 438
Boudegzdame, N. 726
Bouillon, P. 813, 823
Bouzille, G. 342
Bowles, J. 103
Boyer, O. 247
Brađašević, E. 131
Bramkamp (Beckmann), N. 1075
Brännvall, R. 143, 267
Bratsas, C. 433
Breia, P. 337
Brenner, A. 127
Briand, S. 83, 891, 893
Brieden, T. 317
Bringay, S. 561, 773
Brix, T. 947
Brix, T.J. 162
Brogonzoli, L. 846
Brosnan, B. 615
Bruinsma, B. 798
Brunelli, L. 803
Buckow, K. 307
Budu, E. 378
Bülow, C. 207
Burgert, O. 149
Burgun, A. 247, 1037
Busch, D. 927
Butt, Z.A. 783, 861
Buvat, J.-A. 905
Cabalero, I. 891
Cajander, Å. 922
Calafiore, M. 856
Calamo, A. 895
Calcaterra, V. 957
Caldeira, L.L. 1027
Caliesch, R. 586
Candefjord, S. 736
Carli, T. 913
Cato, K. 881, 907
Cecconi, F. 297
Cha, W.C. 651
Chaplinskaya-Sobol, I. 93
Charlet, J. 745, 793
Charon, C. 350
Chasseur, M. 957
Chatellier, G. 202
Chatzimallis, C. 433
Chen, H. 53
Chen, H.H. 783, 861
Chen, X. 1037
Chen, Y.-L. 1067
Chie, E.K. 1015
Chimphoklang, S. 222
Chintiroglou, M. 282
Cho, I. 374
Choi, B. 392

Choi, B.Y. 322
Choi, I. 392
Choi, I.Y. 322
Choi, W. 322, 392
Chondrokostas, E. 433
Christensen, B. 531
Christmann, L.-S. 691
Chytas, A. 551
Cillessen, F. 886
Clausen, S.S. 207
Clemente, C. 337
Cleva, E. 474
Coccia, A. 962, 1029
Coldewey, B. 626, 680
Conceição, G. 492
Conceição, L. 516, 1013
Consorti, F. 506
Coronado, J.M. 871
Corrente, G. 1013
Cort Madsen, I. v
Costa, B. 382
Coulet, A. 768
Coviello, D. 386
Crameri, K. 133
Cruz, Dinis.B. 1071
Cruz-Correia, R. 48, 145
Cuggia, M. 342
Cui, H. 1023
Cui, W. 1023
Cummings, C. 511
D'Addio, G. 962, 1029
D'Uffizi, A. 506
Dagklis, T. 571
Dahlweid, M. 13
Dalainas, I. 576
Damiani, A. 153
Dan, A. 18
Daniel, C. 202
Danilov, G. 972
Darroudi, A. 443
Darwich, A.S. 18
Dathe, H. 977
David, P. 433
Davoody, N. 942
de Boer, M. 348
De Graaf, C. 813
de Hond, A. 817
de Kroon, C. 605
De Petris, L. 18
Debien, C. 474
Dehdeleh, V. 987
Delgado, J. 212, 337
Della Mea, V. 763
DeLong, R. 337
den Otter, S. 139
Denecke, K. 157, 403, 478, 641
Deserno, T.M. 98, 118, 1002
DESIREE Study Group 656
Dettori, S. 380
Dhanasekaran, S. 876
Dhar, E. 153
Dhrangadhariya, A. 586
Di Meco, G. 380
Dickhoner, J. 476
Dieterich, C. 741, 819
Dietrich, M. 937
Dimitriadis, I. 376
Dimitriadis, V. 551
Dimitsaki, S. 396
Dissemond, J. 927
Dittrich, F. 423
Doldi, F. 33
Dorronzoro, E. 403
Douillet, M. 247
Draeger, C. 272
Dreher Sköld, P. 360
Drenkhahn, C. 706
Dryselius, P. 556
Duclos, C. 394
Duftschmid, G. 899
Dugas, M. 137
Dunham, J. 368
Dunn, A. 596
Dunoyer, C. 561, 773
Ebben, K. 605
Eckardt, K.-U. 28
Eckardt, L. 33
Egbert, N. 438
Ehlers, S. 947
Ehrentreich, S. 362
Ehrler, F. 197, 418
Eich, L. 470
Eilbeck, K. 498
Eils, R. 691
Elia, M. 917
Elison-Davies, S. 967
Ellenius, J. 601
Elmahdy, M. 909

Eloranta, A.-M. 1009
Elsig, S. 586
Engel, C. 167
Engelsma, T. 498, 646
Epiphaniou, G. 337
Erdfelder, F. 317
Erfurt-Berge, C. 927
Eriksen, J. 468
Erwee, E. 337
Escalona-Cuaresma, M.J. 386
Esdar, M. 661
Eslami, S. 987
Eter, N. 581, 947
Etminani, K. 352, 378, 556, 609, 613
Fabacher, T. 768
Fanconi, C. 817
Fanelli, V. 566
Fares, E. 474
Farrokhnia, N. 942
Faviez, C. 1037
Faysel, M.A. 511
Fegeler, C. 68, 135, 252
Felicia Faienza, M. 23
Félix, B. 354
Feretzakis, G. 536, 576
Fernandes, A.R. 48
Fernández-Llatas, C. 641
Fetz, B. 803
Figueiredo, B. 382
Filippidis, P.-M. 433
Finger, R. 167
Finkelstein, J. 458, 596, 866, 897, 982, 1023
Fleitmann, M. 952
Flöel, A. 757
Flores, O. 337
Flormann, M. 118
Fluck, J. 108
Fogelberg, E. 556
Forsgren, H. 267
Forster, A. 53
Fortmann, J. 611
Frank, A. 819
Fraser, H. 388
Freeman, C.W. 277
Freitas, A. 492, 516, 1013
Fridgeirsson, E.A. 129, 1057
Friedlander, L. 247
Fruchart, M. 502, 856
Fujarski, M. 33, 237
Fürschke, A. 952
Furukawa, T. 821, 901
Gabarron, E. 403, 478
Gabriel, A. 1023
Gabriel, A.S. 982
Gaitan, S. 1007
Gallos, P. v, 332, 337
Galoyan, T. 476
Ganapathy, N. 73
Ganesan, R. 364
Ganzinger, M. 137, 498
Garcelon, N. 247, 1037
Garin-Muga, A. 521, 1042
Garza, M.Y. 217
Gazzarata, R. 846
GCKD Investigators 28
Gebhardt, M. 307
Gebler, R. 753
Geerts, B. 348
Gehrmann, J. 125
Geis, N.A. 819
Gemke, P. 977
Gencturk, M. 113
George, R. 408
Gerdes, M. 937
Gerlach, J. 823, 952
Gharehbaghi, A. 526
Ghouchan Nezhad Noor Nia, R. 987
Giacobbe, D.R. 380
Giacomini, M. 380, 386, 846
Giacomino, K. 586
Giannios, G. 551
Gierend, K. 741
Gira, M. 778
GMDS Presidential Commission for the Support of Young Scientists 484
Göbel, C. 626
Goris, A. 1007
Gottenberg, J.-E. 768
Govarthan, P.K. 73, 257, 1047
Graessner, H. 292
Grammatikopoulou, M. 551
Greer, M.L. 217
Griga, L. 470
Gruendner, J. 307
Günster, C. 3
Günther, A. 696

Günther, F. 967
Guo, L. 682
Gütebier, L. 747
Haas, P. 1002
Haase, L. 1017
Haavisto, E. 141
Habli, I. 38
Hada, H. 486
Hafer, G. 927
Hager, A. 480
Hägglund, M. v, 242, 356, 942, 1011
Hagiwara, M.A. 736
Hagstedt, L.J. 1011
Hagström, J. 242
Håkansson, A. 876
Halpern, M. 500
Hamed, O. 556, 613
Hammar, T. 207, 287
Hammarlin, M.-M. 798
Hampf, C. 68
Handels, H. 952
Hansen, S. 337
Hanseth, O. 177
Hanss, S. 93, 302
Harkener, S. 167
Harrison, D. 1069
Hartvigsen, G. 841
Hashemi, A.S. 352, 603
Hassankhani, A. 277
Hassmann, J. 661
Haug, M. 755
Havrylchuk, I. 1023
Hawthorne, C. 1069, 1073
Hayn, D. 803
He, M. 819
Hechtel, N. 684
Heggemann, J. 927
Heidemann, G. 927
Heikkilä, A. 617
Heinz, C. 167
Hellmann, F. 917, 932
Hendriks, A. 927
Henke, C. 721
Henke, E. 3, 711, 751
Henke, J. 871
Henkel, R. 147, 747, 749
Henriksen, A. 841, 876
Henselin, K. 871
Henze, R. 118
Her, S. 1015
Hernandez-Boussard, T. 817
Herrmann, C. 438
Hertzum, M. 621
Hess, S. 83
Heuschmann, P. 93
Hilbey, J. 745, 793
Hilfiker, R. 586
Hine, A. 103
Hiragi, S. 486
Hochheim, I. 13
Hofdijk, J. 886
Hoffmann, M. 287, 327
Hoffmann, W. 93
Honka, A. 1009
Hopff, S.M. 93
Horsch, A. 841
Hsieh, T.-C. 932
Huberts, W. 364
Hübner, U. 661, 927
Hund, H. 68, 252
Huo, X. 458, 866
Hüsers, J. 927
Hussein, R. 8, 899
Hustinx, A. 932
Hwan Kim, S. 1015
Iakovidis, D.K. 992
Iardino, R. 846
Ibañez-Sánchez, G. 641
Idrobo-Avila, E. 977
Ihle, R. 317
Ikonen, J. 504
Ikonen, T.S. 413
Illig, T. 93
Im, Y. 392
Im, Y.S. 322
Inau, E.T. 390
Ingenerf, J. 706
Ishizumi, A. 893
Jais, J.-P. 716
Jajroudi, M. 903
Jamalirad, H. 903
Jankovic, M. 202
Jaspers, M. 646
Jaulent, M.-C. 396
Jeleff, A. 418
Jenetzky, E. 167
Jensen, R.A.A. 472
Jensen, V.V. 468

Jersch, P. 167
Jeswani, J. 257
Jiang, K. 833
Jodogne, S. 997
Johann, T. 741
Johannessen, E. 841
Johansson, J. 841
Jonasson, C. 472
Jormanainen, V. 227
Jøsendal, A.V. 207
Jung, H. 78
Jung, S. 322, 392
Junger, D. 149
Kafaie, S. 701
Kahn Jr., C.E. 277
Kaila, M. 227
Kalaeva, D. 972
Kalathe, M. 257
Kalles, D. 536, 576
Kamal, Md.M. 43
Kamonroek, N. 222
Kang, H. 78
Kang, J. 651
Kankainen, K.J.I. 829
Kano, R. 821
Kant, I. 817
Kant, I.M.J. 815
Kapsner, L.A. 58
Karagianni, E. 576
Karanikas, H. 282, 551
Karantanos, P. 576
Karlsen, R. 876
Karlsson Lind, L. 360
Karlsson, T. 433
Karolos, I. 433
Kasbohm, E. 871
Kasiri Habibabadi, S. 98
Katsari, E. 757
Katsch, F. 899
Katzis, K. 337
Kaufman, D. 881, 907
Kempe, N. 470
Kempf, E. 202
Keskimäki, I. 227
Kharko, A. 242, 356
Kim, E.M. 374
Kim, J. 392
Kim, J.Y. 322
Kim, S.J. 651
Kim, S.M. 651
Kim, Y. 1015
Kimiafar, K. 443
Kinnunen, U.-M. 504
Kiriakou, C. 819
Kirsch, E. 149
Kitkowska, A. 666
Kivekäs, E. 504
Klausen, A.D. 656
Klawunn, R. 470, 684
Kleinert, P. 307
Klopfenstein, S.A.I. 133
Knaup, P. 498
Knudsen, C. 262
Ko, S. 392
Ko, S.J. 322
Kobayashi, D. 901
Koch, H. 1025
Kohler, S. 691
Kohn, L. 1019
Koïvogui, A. 394
Kokkinakis, D. 798
Kolde, R. 755, 831
Koledova, E. 23
Kombeiz, A. 362
Kompatsiaris, I. 551
Konakova, T. 972
Korhonen, P. 141
Korntheuer, R. 899
Kors, J.A. 129, 1057
Koufopoulou, C. 536
Kouremenou, E. 153
Kouroumalis, E. 576
Kpenou, F. 716
Kramer, D. 788
Kramer, F. 917, 1075
Kraus, M. 93
Krawitz, P. 932
Krebs, M.-O. 745
Krefting, D. 93, 292, 302, 741, 977
Kreiner, K. 803
Kreuzthaler, M. 788, 825, 827, 837
Kristensen, A.Ø. 468
Kristensson Ekwall, A. 942
Kristoffersen, M.B. 682
Kröner, S. 661
Krückeberg, J. 470, 684
Krüger, E. 757
Kruse, H.M. 317

Kücherer, C. 149
Kugic, A. 827
Kukec, A. 913
Kulau, U. 1031
Kuntz, A. 721
Kuroda, T. 486
Kurscheidt, M. 68, 252
Kurth, A. 366
Kushniruk, A. 907
Kushniruk, A.W. 881
Kutafina, E. 368, 1025
Kuusisto, A. 141
Kyparissidis Kokkinidis, I. 571
Kyriazis, D. 153
Kysh, L. 476
La Mattina, A.A. 364
Lachanoudi, S. 433
Lagakis, P. 376
Laine, H. 617
Lakka, T. 1009
Laleci Erturkmen, G.B. 113
Lame, G. 202
Lamer, A. 474, 502, 851, 856
Lamy, J.-B. 726
Larbi, D. 478
Larburu, N. 521
Larizza, C. 957
Larsson, I. 676, 678
Laukkavirta, M. 413
Laur, S. 831
Lauschenski, A. 803
Lavery, L. 893
Lawin, D. 370, 423
Lawton, T. 38
Le Baher, H. 561
Le Corre, P. 342
Leandro, G.D.S. 172
Leclerc Du Sablon, N. 768
Leddig, T. 871
Lee, E. 736
Lee, H. 372, 1015
Lee, J. 83
Lee, J.-H. 374
Lee, M.S. 78
Lee, S.M. 372
Lei, X. 783
Lemaitre, M. 636, 856
Lewensohn, R. 18
Li, F. 1027
Li, H. 123
Liedes, H. 1009
Liljeberg, E. 398
Lindberg, A. 398
Lindgren, H. 453
Lindgren, M. 227
Lindsköld, L. v
Lindström, J. 1009
Linge, H. 267
Linge, H.M. 143
Linke, F. 1025
Lintu, N. 1009
Lintvedt, O. 192
Lipprandt, M. 611, 626, 656, 680
Listenmaa, I. 829
Liu, H. 1062
Llorente, S. 212, 337
Löbe, M. 272, 721
Loeffler, M. 835
Logaras, E. 376, 571
Lopez Campos, G. 1069, 1073
López, R. 490
Lorbeer, R. 93
Lorenz, S. 753
Lorenz-Depiereux, B. 93
Loupelis, E. 536
Loureiro, M. 48
Lowitsch, V. 366
Lulejian, A. 476
Lundgren, L. 556
Lundholm, T. 731
Lundström, J. 352, 603
Lutze, S. 757
Maag, A. 151
Maalem, S. 636
Maas, J. 1033
Macia, I. 521
Maggi, N. 386
Maglogiannis, I. 332, 337
Maischak, J. 661
Majeed, R.W. 362, 611, 696
Malek, N. 1019
Malihi, L. 927
Maló, P. 337
Mang, J.M. 58
Manias, G. 153
Maniero, A. 763
Männikkö, R. 1009
Manninen, S. 1009

Manskow, U.S. 207
Mantas, J. 337
Maple, C. 337
Marcilly, R. 636
Marco Ruiz, L. 1073
Marco-Ruiz, L. 192
Marie Klöckener, A. 626
Marin, L. 13
Markus, A.F. 1057
Marquardt, K. 358
Marques, R. 354
Marreiros, G. 516, 1013
Martens, U.M. 135
Martin, D. 167
Martin, T. 292
Martinez, M. 1042
Martinez-García, A. 386
Martinez-Herrera, M.R. 521
Martînez-Millana, A. 641
Martins, R. 1013
Martinuzzi, A. 763
Marx, G. 366
Marzano, L. 18
Master Östlund, C. 496
Matera, R. 895
Mateus, M. 48
Matragkas, N. 337
Mattay, R.R. 277
Mattila, E. 1009
Mayer, S. 1075
McEvoy, E. 1069
Meijer, S. 18
Meineke, F. 835
Meinke, F. 721
Melander, O. 609
Melo, R. 354
Melot, B. 905
Menon, R.N. 232
Menychtas, A. 332, 337
Menzel, F. 749
Mertes, S. 917
Meyers, D. 317
Meystre, S. 1007
Michaelis, L. 147, 741
Michel-Backofen, A. 358
Miletic, M. 641
Miller, T. 511
Minshall, S. 881
Misawa, S. 821
Mlakar, A. 913
Modersohn, L. 835
Mödinger, S. 68, 252
Moelleken, M. 927
Moen, A. v
Moen, H. 344
Mohammadi, F.G. 911
Molina, C. 546
Molinnus, D. 366
Monti, F. 905
Moon, S.H. 651
Mora, S. 380
Mori, Y. 486
Moro, C. 172
Moro, C.H.C. 172
Moros, L. 773
Mosteiro, P.J. 815
Mouazer, A. 726
Mousavi Baigi, S.F. 443
Mraidha, C. 337
Mueller, A. 1019
Mueller, D. 917
Mujtaba, M.M. 833
Mukherjee, S. 1047
Muli, I. 1011
Müller, A. 691
Müller, D. 1075
Müller, H. 586
Müller, J. 470
Muñoz Monjas, A. 88
Murgia, Y. 846
Muzoora, M.R. 741
Naderi, N. 586
Nagel, V. 172
Nair, M. 556
Nakkas, H. 388
Namer, B. 368
Natarajan, K. 312
Natsiavas, P. 384, 396, 551, 686
Neglia, G. 895
Neuraz, A. 247
Neururer, S. 803
Névéol, A. 768
Nguyen, T. 83, 891, 893
Ni, Z. 123
Niebauer, J. 8
Niekrenz, L. 611
Nielsen, C.P. 631
Nielsen, J. 631

Nielsen, K. 472
Nielsen, M.S. 1021
Niklas, C. 137, 747
Nikolaidis, G. 551
Nikolopoulos, S. 551
Nilsson, I. 453
Nilsson, L. v, 287
Nøhr, C. 428, 448
Noll, R. 607
Noor, H. 53
Nörby, U. 360
Norré, M. 823
Nøst, T.H. 1021
Notredame, C.-E. 474
Nour, R. 463
Novoa, M.M. 682
Nsais, T. 905
Nyberg, M. 398
Nygren, J. 346, 556
Nygren, J.M. 676, 678
Oh, S. 392
Oh, S.W. 322
Ohlsson, M. 609
Oja, M. 755
Oldenburg, J. 8
Olender, S. 500
Oliveira, J.L. 1071
Oliveira, M. 48
Oluka, H. 881
Onoda, K. 821
Orrevall, Y. 398
Ortiz-Catalan, M. 682
Ottmann, I. 684
Oudbier, S.J. 498
Oyama, S. 821, 901
Ozturk, B. 38
Padilla, J.J. 500
Pagano, G. 962, 1029
Pajusalu, M. 755
Palchuk, M. 88
Pallenberg, R. 952
Pålsson, P. 496
Panagopoulos, C. 332, 337
Panoutsakou, M. 576
Panteris, V. 576
Paolucci, M. 297
Papachristou, P. 337
Papalois, A.E. 576
Paparella, G. 763
Paraskevopoulos, E. 1052
Pardos, A. 332
Park, H.-A. 78
Park, S. 392
Park, S.K. 78
Park, Y. 372, 372
Parra-Calderón, C.L. 386
Partovi, E. 526
Parvanova, I. 458, 596, 982, 1023
Paschen, S. 757
Pastorino, A. 891
Patino-Martínez, M. 153
Pecoraro, F. 297, 506
Pedersen, E.R. 631
Pedersen, R. 192
Peltonen, L.-M. 344, 617
Peng, Y. 711, 751, 1062
Pérez Garriga, A. 611
Pérez-Rey, D. 88
Petersson, L. 676, 678
Petersson, M. 556
Petković, M. 1033, 1033
Peute, L. 646
Peute, L.W. 498
Pfeifer, B. 788, 803, 827
Pichon, A. 8
Pickering, J. 433
Pierrakou, A. 576
Pierre-Jean, M. 342
Pihlajamäki, J. 1009
Piho, G. 829
Pilgram, L. 28
Pinheiro, V. 382
Pinto, A. 492
Plagwitz, L. 33, 127, 182, 237
Pliatsika, J. 551
Pölzl, G. 803
Poncelet, P. 561, 773
Popova, A. 972
Porschen, C. 237
Postnov, A. 972
Powell, L. 463
Poyraz, R.A. 741
Prados-Suarez, B. 546
Pragya 1047
Prasser, F. 28, 691
Premanandan, S. 922
Prince, A. 257
Priou, S. 202

Prokosch, H.-U. 58, 307
Pronin, I. 972
Przysucha, M. 927
Purnat, T.D. 891, 893
Pustozerov, E. 370, 423
Qian, Y. 123
Quaglini, S. 957
Quast, K.-J. 317
Quennelle, S. 247
Quiles, J. 490
Quindroit, P. 636, 851, 856
Quirant, L. 601
Raboudi, A. 745
Radke, D. 871
Raghothama, J. 18
Rahme, J. 591
Rajabi, E. 701
Rajamanickam, Y. 232
Ramirez, S.O. 500
Ranegger, R. 151
Ranjbar, A. 177, 488
Ratz, M. 337
Rausch, D. 13
Ravn, J. 177, 488
Rb-Silva, R. 516
Rebescher, K. 521
Reichenpfader, D. 403, 778
Reif, W. 917
Reinecke, I. 3, 711, 753
Reineke, A. 871
Rejc, T. 913
Reps, J.M. 129, 139
Rian, O. 472
Ribeiro, X. 516
Ricci, F.L. 506
Richter, A. 871
Richter, M.L. 927
Richter-Pechanski, P. 819
Rigas, E.S. 571
Rijnbeek, P.R. 129, 139, 1057
Rivera Romero, O. 641
Rivera-Romero, O. 23
Robel, L. 716
Rodrigues, D. 1013
Rodriguez, N. 561
Rögnvaldsson, T. 378
Röhrig, R. 362, 368, 611, 626, 656, 680, 1025
Roitero, K. 763
Roller, R. 825
Roloff, M. 757
Ronningen, E. 177
Rosati, S. 566
Rosenau, L. 307
Rosón, B. 490
Ross, P. 829
Rossander, A. 671
Rostam Niakan Kalhori, S. 98
Rotegård, A.K. 531
Roth, H. 757
Rothermel, M. 1025
Roy, M. 817
Rubio Ruiz, D. 88
Rubio-Escudero, C. 521
Ruhnke, T. 3
Ruotsalainen, I. 1009
Rupp, R. 167
Saab, A. 591
Sadjadi, M. 237
Saint-Dizier, C. 474, 851
Saito, K. 486
Sakagianni, A. 536
Salgado, P.F. 354
Salzano, A. 895
Samo, G. 743
Sampaio, S. 1013
Sample, S. 217
Sandberg, L. 601
Sandsjö, L. 682
Santos, C. 492
Santos, P. 492
Santos, R.L. 145
Saranto, K. 141, 504
Sarbaz, M. 443
Sareban, M. 8
Sarris, A.L. 1052
Sasse, J. 108
Sato, K. 821, 901
Sattelmayer, M. 586
Sauleau, E.-A. 768
Saunier, S. 1037
Savage, C. 346
Savino, M. 153
Sax, U. 721
Scandurra, I. 242
Schaaf, J. 607
Schäfer, C. 93
Schäffner, E. 28

Schaller, J. 93
Schauer, B. 871
Scheble, V. 1019
Schiesser, L. 581
Schießer, L. 947
Schinkelshoek, L. 348
Schinle, M. 937
Schiza, E. 433
Schlebusch, F. 368
Schmidt, C.O. 272, 871
Schmidt, T. 448
Schnall, R. 500
Schneider, B. 438
Schneider, H. 358
Schoenthaler, M. 167
Scholz, M. 871
Schöneck, M. 1027
Schöner, M. 757
Schössow, J. 871
Schreier, G. 803
Schreiweis, B. 541
Schriewer, E. 1025
Schroeter, N. 917
Schulz, S. 788, 825, 827, 837
Schumacher, D. 317
Schüttler, C. 484
Schwab, D.M. 819
Schwab, U. 1009
Schweizer, S.T. 68, 252
Scott, P. 388
Scotto, A. 566
Sebro, R. 909, 911
Sedki, K. 726
Sedlmayr, M. 3, 711, 751, 753
Segura, M. 641
Sellemann, B. 438
Selvaraj, A. 257
Semler, S.C. 307
Senathirajah, Y. 881, 907
Sengupta, S. 312
Seo, Y. 651
Seroussi, B. 591
Serth, S. 408
Servajean, M. 773
Severinsen, G.H. 187
Shabbir, S.A. 153
Shah-Mohammadi, F. 897
Shaik Gadda, A.A. 232
Sharma, K. 53
Sheridan, K. 83
Shiferaw, K.B. 63
Shifrin, M. 972
Shiratori, Y. 821, 901
Shugay, S. 972
Sidiropoulos, E. 433, 1052
Sigle, S. 135
Siira, E. 346
Siladlao, S. 222
Silsand, L. 187
Silva, J.M. 1071
Singer, J. 511
Singh Dhillon, N. 13
Sinha, K. 1047
SITA (Italian Society of Anti-Infective Therapy) 380
Sjöqvist, B.A. 736
Skarga-Bandurova, I. 615
Skarha-Bandurov, I. 615
Skolt, K. 488
Sletvold Øistad, B. 488
Smed, J. 617
Smiley, A. 1023
Smith, S. 38
Soares, M. 337
Sobotta, V.M.G. 118
Söderholm, H.M. 736
Soleiman, J. 98
Soliman, A. 352, 378, 556, 613
Son, M.H. 651
Sørensen, C.B. 631
Sørensen, T.H. 468
Souza, J. 516, 1013
Sovatzidi, G. 992
Spechbach, H. 813, 823
Speranza, M. 716
Spicher, N. 977
Spoden, M. 3
Spreckelsen, C. 438
Sreeraj V S 257
Sriram Kumar P 73
Stahl, D. 93
Stamkopoulos, E. 376
Stasiak, L. 337
Stathakarou, N. 433
Staubitz, T. 408
Stausberg, J. 167
Stecca, G. 506
Steinsbekk, A. 187, 1021

Stellmach, C. 133
Steyerberg, E. 348, 817
Stock, S. 937
Stockem, C. 611
Stöhr, M.R. 696
Storck, M. 162, 947
Storf, H. 607
Stork, W. 937
Storp, J.J. 581
Storz, P. 1019
Stracke, S. 757
Strahwald, B. 438
Stroth, A.M. 952
Stroth, D. 237
Struckmann, S. 871
Subramaniam, V. 257
Sucre, A. 1042
Sümer, Ö. 932
Sumunar, D. 433
Sung, S. 78, 1015
Šuvalov, H. 831
Suwelack, B. 167
Svedberg, P. 346, 676, 678
Tagaras, K. 433
Tahar, K. 292
Taloyan, M. 1011
Tamburis, O. 895
Taniguchi, T. 821
Tarkkala, K. 413
Tasoulis, S. 282
Tatiyaworawattanakul, K.H. 222
Taveira-Gomes, T. 516
TEDIS Study Group 716
Tendler, S. 18
Theodorakopoulou, M. 576
Thies, C. 1019
Thomaidis, T. 576
Thomas, J. 232, 257
Thun, S. 133, 741
Thurin, A. 759
Tienari, M. 413
Tilles-Tirkkonen, T. 1009
Timonen, J. 207
Tiwari, P. 603
Toddenroth, D. 808
Tokar, A. 408
Tombolini, G. 962, 1029
Tooulias, A. 433
Töpfer, P. 757
Tournigand, C. 202
Troglio, A. 368
Truong, T.-T. 596
Tsai, T.-Y. 1023
Tsakiridis, I. 571
Tsao, S.-F. 53, 783, 861
Tsioukas, V. 433
Tsopra, R. 726
Tsoulfas, G. 433
Tudpor, K. 222
Turnbull, N. 222
Turner, P. 428
Tute, E. 272
Ulas, S.B. 1019
Ulrich, H. 484
Umer, A. 1009
Utunen, H. 408
Uva, P. 386
Vakali, A. 376
Valentin, H. 93
Valero-Ramón, Z. 641
Vallo Hult, H. 496
Van Alen, C.M. 127, 237
van Boekel, A. 348
van Buchem, M. 817
van Buchem, M.M. 815
van de Vosse, F.N. 364
van der Meijden, S.L. 348
van der Werf, J. 605
Van Gils, M. 1009
van Goor, H. 348
van Houdt, M. 605
van Stiphout, R. 1007
Varella, A. 433
Varghese, J. 33, 127, 162, 182, 237, 438, 581, 947
Vedantham, D. 232
Vehko, T. 227
Vehreschild, J. 93
Venturi, A. 846
Verbuecheln, R. 292
Verdier, L. 502
Vergani, E. 23
Verhamme, K.M.C. 1057
Verykios, V. 536
Verykios, V.V. 576
Vgenopoulou, S. 576
Viana, J. 492
Vidal, A.C. 337

Vieira, A. 1013
Vikhrova, N. 972
Villumsen, S. 472
Vincent, K. 676
Vincent, M. 1037
Vistocco, D. 895
Vogelsang, T. 33
Volmerg, J.S. 680
Völzke, H. 871
Von Gerich, H. 344
Von Groote, T. 237
Von Jan, U. 370, 423
von Kaenel, F. 641
Von Wagner, M. 607
Voulgarakis, V. 433
Vuorinen, A.-L. 1009
Waltemath, D. 63, 272, 390, 741, 747, 749, 757
Wannheden, C. 480
Webber, T. 103
Weber, P. 337
Weber, Y. 611, 1025
Wegner, J. 167
Weiss, R. 237
Weng, C. 312, 1062
Wermundsen Mork, E. 488
Werner, P. 135
Westphal, S. 871
Wettstein, R. 68
Whatelet, M. 474
White, B. 891, 893
Wiedekopf, J. 706
Wiemeyer, S. 927
Wiesenbach, P. 819
Wiesmüller, F. 803
Williams, R.D. 129, 139
Williamson, R. 337
Winkelhake, L. 470
Winter, A. 272, 438, 721
Wodke, J.A.H. 147
Wolking, S. 611, 1025
Wong, D. 967
Wong, J. 129
Wong, M.C. 428
Wright, A. 893
Wuillemin, P.-H. 350
Wynn, R. 478
Xavier, W. 1013
Xhakli, T. 1023
Xing, Z. 596
Yamashita, S. 901
Yang, C. 129
Yang, Y. 783, 861
Yao, Z. 123
Yarimizu, H. 821
Yasini, M. 13
Yasuoka, M. 621
Yau, B. 83, 893
Yau, C. 967
Yee, K.C. 428
Yildirim, K. 581, 947
Yoon, H.-J. 372
You, H. 372
You, M. 1015
Yusuf, K.O. 302
Zaanouar, M. 851
Zachariadou, M. 551
Zagorodnova, A. 972
Zakashansky, E. 1023
Zanaboni, P. 478
Zarbock, A. 237
Zary, N. 463
Zecchin, T. 893
Zeleke, A. 63, 390
Zenker, S. 317
Zhao, Y. (Ursula) 408
Zhou, Z. 783
Zilske, C. 68, 252
Zoch, M. 3
Zoubouli, C. 576
Zucchiatti, C. 23
Zuccotti, G. 957

This page intentionally left blank