Predicting Exacerbations in Patients with Chronic Obstructive Pulmonary Disease

Kronborg, Thomas

Publication date:
2019

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

PREDICTING EXACERBATIONS IN PATIENTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE

BY

THOMAS KRONBORG LARSEN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY

DENMARK

PREDICTING EXACERBATIONS IN PATIENTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE

by

Thomas Kronborg Larsen

AALBORG UNIVERSITY
DENMARK

Dissertation submitted December 2019

Dissertation submitted:	December 19th 2019
PhD supervisor:	Professor Ole Hejlesen Aalborg University
Assistant PhD supervisor:	Postdoc Simon Lebech Cichosz Aalborg University
PhD committee:	Associate Professor Lasse Riis Østergaard Aalborg University
	Associate Professor Christian Fischer Pedersen Aarhus University
	Associate Professor Raúl San José Estépar Harvard Medical School
PhD Series:	Faculty of Medicine, Aalborg University
Department:	Department of Health Science and Technology

ISSN (online): 2246-1302
ISBN (online): 978-87-7210-566-6

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

CV

Thomas Kronborg Larsen completed his master's degree in Biomedical Engineering and Informatics at Aalborg University in 2015. As a continuation of his master's thesis, Thomas worked as a research assistant on developing models for predicting exacerbations in patients with chronic obstructive pulmonary disease (COPD) based on data from a large-scale telemedicine trial conducted in the North Denmark Region. In August 2016, Thomas was enrolled as a PhD student at the Doctoral School in Medicine, Biomedical Science and Technology at Aalborg University.

His PhD study evolved around improving models for predicting exacerbations in patients with COPD under supervision by Professor Ole Hejlesen and co-supervision by Postdoc Simon Lebech Cichosz. During his studies, Thomas acquired the necessary knowledge in statistical models and classification algorithms to develop predictive models based on machine learning and artificial intelligence. In the spring of 2016, Thomas was awarded and completed a four-month Fulbright scholarship at the University of Washington, Seattle, Washington, USA. His work has led to journal publications and presentations at Medical Informatics Europe and at European Respiratory Society's International Congress.

Thomas has also participated in several activities alongside his PhD study. He has conducted several clinical studies related to COPD at Aalborg University and North Denmark Regional Hospital. This work has also led to publications and to Aalborg University taking two patents. He has also been highly engaged in teaching, supervising ten semester projects among Biomedical Engineering and Informatics, Clinical Science and Technology, and Medicine. Recently, Thomas has also taken up case supervision in Medicine and Medicine with Industrial Specialization.

ENGLISH SUMMARY

Chronic obstructive pulmonary disease (COPD) is a common, preventable, and treatable disease characterized by persistent respiratory symptoms and airflow limitation. The disease usually arises due to airway and/or alveolar abnormalities that are often caused by significant exposure to noxious particles or gases. COPD is a major socio-economic burden that is associated with significant morbidity and mortality worldwide. The burden of the disease is mainly due to exacerbations, which are defined as acute worsenings of respiratory symptoms that result in additional therapy. COPD exacerbations have a negative impact on health status, rate of hospitalization and readmission, and disease progression. Although exacerbations are both treatable and preventable, a substantial portion of unreported exacerbations is a challenge in the management of COPD. Telemonitoring of patients with COPD and the application of predictive analytics have been proposed as a potential solution, and several previous studies have attempted to develop predictive models for upcoming exacerbations. However, no previously reported studies have achieved a level of clinical relevance for implementation in the telemonitoring of patients with COPD. Although previous exacerbations are clinically recognized as a predictor of future exacerbations, no previous studies have investigated this potential.

The objective of this thesis was to improve methods for assessing predictive models and to explore the potential of including information on previous exacerbations in the prediction of upcoming exacerbations in patients with COPD. The thesis was based on four studies presented in four individual papers. Study I and study IV explored the potential of including information on previous exacerbations, whereas study II and study III involved methodological research for assessing predictive models leading up to study IV. Study I investigated whether patient and population exacerbation incidence contain predictive information of upcoming exacerbations. This study was based on home measurements of physiological parameters in a large cohort of telemonitored patients with COPD, but the study was limited by the frequency of the submitted measurements. To determine whether the predictive models could be improved with a higher frequency of measurements, a smaller cohort of telemonitored patients with more frequent measurements was investigated in studies II-IV. A challenge when developing predictive models in a smaller cohort is over-fitting. Therefore, study II investigated whether double cross-validation with feature selection was a more robust validation method than conventional cross-validation for preventing over-fitting and ensuring generalizable test results. Study III systematically compared nine classification algorithms to assess the sensitivity to the choice of classification algorithm for predictive models for COPD exacerbations.

Study IV combined the work of studies I-III and investigated whether a two-layer probabilistic model could increase classification rates compared to conventional models for predicting COPD exacerbations.

The conducted studies revealed that the population exacerbation incidence holds predictive information of upcoming exacerbations for telemonitored patients. Double cross-validation with feature selection proved to be a more robust validation method than conventional cross-validation for predicting COPD exacerbations in a small imbalanced data set. Prediction models for COPD exacerbations are highly sensitive to the choice of classification algorithm, and several classification algorithms should therefore be compared systematically when developing predictive models. Finally, a two-layer probabilistic model can significantly increase the classification rates to a level of clinical relevance compared to a conventional one-layer model for predicting COPD exacerbations. In conclusion, double cross-validation should be used to develop and compare models based on several classification algorithms, and there is great potential in including information on previous exacerbations in the prediction of upcoming exacerbations in patients with COPD. This could allow for clinical implementation and potentially reduce the morbidity, mortality, and socio-economic burden related to COPD.

DANSK RESUME

Kronisk obstruktiv lungesygdom (KOL) er en almindelig, forebyggelig og behandlelig sygdom, der er kendetegnet ved vedvarende luftvejssymptomer og luftvejsobstruktion. Sygdommen skyldes normalt luftvejs- og/eller alveolære abnormaliteter, som oftest forårsages af en betydelig eksponering for skadelige partikler eller gasser. KOL er en stor socioøkonomisk burde, som er associeret med betydelig sygelighed og dødelighed verden over. Sygdomsbyrden skyldes primært exacerbationer, der defineres som akutte forværringer af luftvejssymptomer, som kræver yderligere behandling. KOL exacerbationer har en negativ påvirkning på sundhedsstatus, indlæggelses- og genindlæggelsesrate og sygdomsprognose. Selvom exacerbationer både kan behandles og forebygges, er en betydelig mængde urapporterede exacerbationer en udfordring i håndteringen af KOL. Telemonitorering af patienter med KOL og anvendelsen af prædiktive analyser er blevet foreslået som en potentiel løsning, hvor flere tidligere studier har forsøgt at udvikle prædiktive modeller for kommende exacerbationer. Imidlertid har ingen tidligere rapporterede studier opnået et niveau af klinisk relevans til at kunne implementeres i telemonitoreringen af patienter med KOL. Selvom tidligere exacerbationer er en klinisk anerkendt prædiktor for kommende exacerbationer, har ingen tidligere studier undersøgt dette potentiale.

Formålet med denne afhandling var at forbedre metoder til at vurdere prædiktive modeller og at undersøge potentialet i at inkludere information om tidligere exacerbationer i prædiktionen af kommende exacerbationer hos patienter med KOL. Afhandlingen var baseret på fire studier, som er præsenteret i fire individuelle artikler. Studie I og studie IV undersøgte potentialet i at inkludere information om tidligere exacerbationer, mens studie II og studie III omhandlede metodisk forskning, der ledte frem til studie IV. Studie I undersøgte, om patient- og populationsincidensen af exacerbationer indeholder prædiktiv information om kommende exacerbationer. Dette studie var baseret på hjemmemålinger af fysiologiske parametre i en stor kohorte af telemonitorerede patienter med KOL, men studiet var begrænset af frekvensen af de indsendte målinger. En mindre kohorte med hyppigere målinger blev undersøgt i studierne II-IV med henblik på at afgøre, om de prædiktivere modeller kunne forbedres med en højere målefrekvens. Over-fitting er en udfordring, når der udvikles prædiktive modeller i en mindre kohorte. Derfor undersøgte studie II, om dobbelt krydsvalidering med valg af prædiktorer er en mere robust valideringsmetode end konventionel krydsvalidering til at forhindre over-fitting og sikre generaliserbare testresultater. Studie III foretog en systematisk sammenligning af ni klassifikationsalgoritmer for at vurdere sensitiviten over for valg af

klassifikationsalgoritme for prædiktive modeller for exacerbationer af KOL. Studie IV kombinerede studierne I-III og undersøgte, om en to-lags probabilistisk model kunne øge klassifikationsraterne, sammenlignet med konventionelle modeller, til prædiktion af exacerbationer af KOL.

De gennemførte studier viste, at populationsincidencen af exacerbationer indeholder prædiktiv information om kommende exacerbationer hos telemonitorerede patienter. Dobbelt krydsvalidering med valg af prædiktorer blev anset som en mere robust valideringsmetode end konventionel krydsvalidering for prædiktion af KOL exacerbationer i et lille ubalanceret data sæt. Prædiktive modeller er yderst sensitive over for valg af klassifikationsalgoritme, og derfor bør flere klassifikationsalgoritmer sammenlignes systematisk under udviklingen af en prædiktiv model. Endelig kan en to-lags probabilistisk model øge klassifikationsraterne signifikant til et niveau af klinisk relevans sammenlignet med en konventionel et-lags model for prædiktion af exacerbationer af KOL. Som konklusion bør dobbelt krydsvalidering anvendes til at udvikle og sammenligne modeller baseret på flere klassifikationalgoritmer, og der er et stort potentiale i at inkludere information om tidligere exacerbationer i prædiktionen af kommende exacerbationer hos patienter med KOL. Dette kunne muliggøre en klinisk implementering og potentielt reducere sygeligheden, dødeligheden og den socioøkonomiske byrde relateret til KOL.

PREFACE

This PhD thesis has been submitted to the Doctoral School in Medicine, Biomedical Science and Technology at Aalborg University in fulfilment of the requirements for the PhD degree. The thesis was based on the scientific work that was accomplished during the PhD study from August 2016 to November 2019 at the Department of Health Science and Technology. The PhD study was supervised by Professor Ole Hejlesen and co-supervised by Postdoc Simon Lebech Cichosz at Aalborg University.

The thesis is based on four studies that were conducted during the PhD study, which are presented in four different papers. The thesis additionally includes a background that presents the research area, which leads to the objective of the thesis. The four studies are then presented along with their individual contributions. The thesis ends in a discussion of the included studies, the significance of the conducted research, and future perspectives along with a conclusion that summarizes the main findings of the thesis.

ACKNOWLEDGEMENTS

First, I would like to express my most sincere gratitude towards my supervisor Ole Hejlesen, who introduced me to research, guided me through my ups and downs in the past three years, and inspired me throughout my PhD study. My thanks and appreciation also go to my assistant supervisor Simon Lebech Cichosz for sharing his excellent technical knowledge and for discussions that have refined my research.

I am very grateful for all my colleagues at the Department of Health Science and Technology, Aalborg University, especially for my friend and colleague, Stine Hangaard Casper, who has endured endless complaining, yet has always remained positive and supportive in our office and throughout our research collaborations.

I would also like to thank my family and my friends for all their interest and support. Last but not least, I would like to express my gratitude towards Stefanie and my son Bjørn for their love and endless support and for inspiring me to always be my very best.

PUBLICATION LIST

THESIS PUBLICATIONS

Paper I:

Kronborg, T., Mark, L., Cichosz, S. L., Secher, P. H., and Hejlesen, O.

Population exacerbation incidence contains predictive information of acute exacerbations in patients with chronic obstructive pulmonary disease in telecare

Published in International journal of Medical Informatics, vol. 111, p. 72-76, 2018

doi: 10.1016/j.ijmedinf.2017.12.026

Paper II:

Kronborg, T., Cichosz, S. L., and Hejlesen, O.

Double cross-validation with feature selection: An overfit-resistant and transparent validation method in a small imbalanced data set for predicting exacerbations in patients with COPD

Submitted to Computers in Biology and Medicine, 2019

Paper III:

Kronborg, T., Cichosz, S. L., and Hejlesen, O.

A systematic approach to sensitivity-analysis of choice of classification algorithm for prediction of COPD exacerbations

Submitted to Computers in Biology and Medicine, 2019

Paper IV:

Kronborg, T., Cichosz, S. L., and Hejlesen, O.

A two-layer probabilistic model shows promising results for predicting COPD exacerbations for patients in telemonitoring

Submitted to International Journal of Medical Informatics, 2019

RELATED PUBLICATIONS

Hangaard, S., Kronborg, T., Stausholm, M. N., Cichosz, S. L., and Hejlesen, O. K.
Using the pre-bronchodilator spirometry curvature to improve estimation of post-bronchodilator airflow obstruction
European Respiratory Society's International Congress, Madrid, Spain, 2019

Hangaard, S., Kronborg, T., and Hejlesen, O. K.
Characteristics of subjects with undiagnosed COPD based on post-bronchodilator spirometry data
Respiratory Care, vol. 64, issue 1, p. 63-70, 2019
doi: 10.4187/respcare.06296

Cichosz, S. L., Stausholm, M. N., Kronborg, T., Vestergaard, P., and Hejlesen, O.
How to use blockchain for diabetes health care data and access management: An operational concept
Journal of Diabetes Science and Technology, vol. 13, issue 2, p. 248-253, 2019
doi: 10.1177/1932296818790281

Kronborg, T., Hangaard, S., Cichosz, S. L., and Hejlesen, O.
Increased Accuracy After Adjustment of Spirometry Threshold for Diagnosing COPD Based on Pre-Bronchodilator FEV1/FVC
Respiratory Care, vol. 64, issue 1, p. 85-90, 2019
doi: 10.4187/respcare.06148

Kronborg, T., Cichosz, S. L., and Hejlesen, O.
Prediction of COPD Exacerbations from Relative Day-To-Day Variations in Physiological Parameters
Proceedings of Medical Informatics Europe, Gothenburg, Sweden, 2018

Hangaard, S., Kronborg, T., and Hejlesen, O.
Feno: a potential biomarker in COPD?
Proceedings of Medical Informatics Europe, Gothenburg, Sweden, 2018

Cichosz, S. L., Kronborg, T. and Hejlesen, O.

Stapedius Reflex and Hearing loss Investigation in People with COPD

Researchgate, 2018

doi: 10.13140/RG.2.2.12867.66082

ABBREVIATIONS

AUC: Area under the curve

CART: Classification and regression trees (machine learning algorithm)

CAT: COPD assessment test

COPD: Chronic obstructive pulmonary disease

FEV_1: Forced expiratory volume in 1 second

FVC: Forced vital capacity

GOLD: Global Initiative for Chronic Obstructive Lung Disease

KOL: Kronisk obstruktiv lungesygdom (Danish)

mMRC: Modified British Medical Research Council dyspnoea scale

ROC: Receiver operating characteristic

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

The following chapter presents the rationale for this PhD study.

Chronic obstructive pulmonary disease (COPD) is a major public health problem that is associated with a substantial impact on morbidity, mortality, and socio-economic costs worldwide. The burden related to COPD is primarily attributable to exacerbations, which are acute worsenings of respiratory symptoms that result in additional therapy. Although exacerbations are both treatable and preventable, unreported exacerbations are a challenge in the management of COPD. Telemonitoring and the application of predictive analytics have been proposed as a potential solution, but there is an urgent need to develop better predictive models for exacerbations. This thesis contributes to the field of COPD research by exploring a new potential to improve the state-of-the-art in the use of predictive analytics for telemonitoring of patients with COPD. The conducted research could potentially lead to clinical implementation and assist patients and healthcare personnel in better managing COPD.

CHAPTER 2. BACKGROUND

This chapter will describe the human respiratory system and the development of COPD. Subsequently, the focus will shift to exacerbations when describing the global burden of the disease and the challenges in the management thereof. Finally, the chapter presents the state-of-the-art in the management of the disease with predictive analytics and what has yet to be explored.

2.1 THE HUMAN RESPIRATORY SYSTEM

The human respiratory system is responsible for taking in oxygen and disposing carbon dioxide within the lungs. The upper respiratory system consists of the nasal cavity, paranasal sinuses, pharynx (throat), and larynx (voice box) [1]. These pathways filter and humidify the incoming air to protect the more delicate surfaces of the lower respiratory system [2]. In the lower respiratory system, the air continues through the trachea (windpipe), which bifurcates into the left and right primary bronchi [1]. The right primary bronchus supplies the right lung, and the left primary bronchus supplies the left lung (Figure 1).

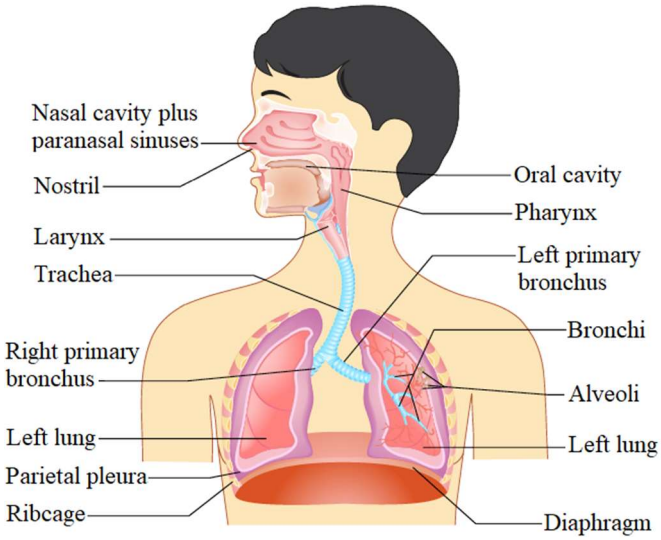

Figure 1. Anatomy of the respiratory system.

The primary bronchi then branch to form secondary and tertiary bronchi. The trachea and main bronchi contain C-shaped rings of cartilage, which serve to support and protect the airways [2]. Each bronchus further divides repeatedly, where each

generation approximately doubles in number, the cartilage progressively decreases, and the amount of smooth muscle tissue increases. Bronchioles begin at approximately generation 12, in which cartilage is absent and the airways are embedded in the elastic lung tissue. Terminal bronchioles (generation 16) lead to respiratory bronchioles, which are the first generation to have alveoli in their walls [1]. Alveoli are air-filled pockets where gas exchange takes place between the lungs and blood [2]. Respiratory bronchioles lead to alveolar ducts and alveolar sacs (generation 23), whose walls are entirely composed of alveoli [1]. A thin epithelial membrane between the alveoli and surrounding capillaries, generally narrower than 1 μm, ensures a quick and efficient gas exchange. To meet the metabolic requirements of the body, the lungs must provide a large surface area for gas exchange. Each lung contains approximately 150 million alveoli, which are associated with an extensive network of alveolar capillaries. The surface area involved in gas exchange has been estimated to range from 70 m^2 to 140 m^2 [2]. During inhalation, skeletal muscles cause an elevation of the ribcage along with a contraction and downward motion of the diaphragm. The increased thoracic volume causes a decrease in the intrapleural and alveolar pressure that draws air into the lungs [1].

The airways from the trachea to the respiratory bronchioles are lined with ciliated columnar or cuboidal epithelial cells. Mucus cells and mucus glands within the epithelium secrete a sticky mucus that lines the exposed surfaces to trap debris or pathogens that have entered the airways. Synchronous movement of the cilia moves the mucus from the lungs towards the pharynx, from where the mucus is swallowed and exposed to the acids and enzymes of the stomach [2].

2.1.1 GAS EXCHANGE

Oxygen and carbon dioxide are exchanged by diffusion though the alveolar-capillary membrane. This gas exchange is a response to concentration gradients from differing partial pressures within the alveolar air and the blood [1]. Inhaled air is a mixture of gases, where nitrogen accounts for 78.6%, and therefore the vast majority of molecules, while oxygen makes up approximately 20.9% percent of the air. The remaining 0.5% of molecules mostly consist of water, with carbon dioxide contributing 0.04%. The combined effect of the molecules in the air results in an atmospheric pressure of 760 mmHg. The partial pressure of oxygen in the dry inhaled air is therefore 159 mmHg, while the partial pressure of carbon dioxide is 0.3 mmHg. As the inhaled air is humidified in the upper respiratory system, the partial pressure of oxygen in the alveoli is 100 mmHg. Upon reaching the alveoli, the inhaled air also mixes with air remaining from the previous exhalation, resulting in a partial pressure of 40 mmHg for carbon dioxide [2].

Deoxygenated blood reaches the alveolar capillaries from the heart though the left and right pulmonary arteries. The blood contains a lower partial pressure of 40 mmHg for oxygen and a higher partial pressure of 45 mmHg for carbon dioxide compared to the alveolar air. This results in a pressure gradient that forces oxygen from the alveoli into the blood and forces carbon dioxide from the blood into the alveoli. The blood then reaches equilibrium with the alveolar air at a partial pressure of 100 mmHg for oxygen and 40 mmHg for carbon dioxide. The oxygenated blood then departs from the alveoli and returns to the heart through the pulmonary veins before it is pumped into the systemic circuit to supply the body with oxygen [2]. An important note is that the body needs more oxygen and generates more carbon dioxide than what can be dissolved in the blood. Red blood cells therefore play an important role in binding molecules dissolved in the blood, which causes a continuous diffusion of gasses [1]. Only 1.5% of the oxygen molecules are dissolved in the blood leaving the alveoli, whereas the rest are bound to haemoglobin within the red blood cells. Carbon dioxide is generated in peripheral tissues and transported either as carbonic acid (~70%), bound to haemoglobin (~23%), or dissolved in the blood (~7%). These reactions are both temporary and reversible, as red blood cells remove excess molecules when the concentrations in the blood are high and release their stored reserve when concentrations are low [2].

2.2 CHRONIC OBSTRUCTIVE PULMONARY DISEASE

COPD is defined as a common, preventable and treatable disease that is characterized by persistent airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases [3]. Long-term exposure causes chronic inflammation in the airways, usually referred to as chronic bronchitis, and causes both small airway diseases and parenchymal destruction. Chronic bronchitis causes a narrowing of the small airways due to mucosal inflammation and secretion and decreases the elasticity of the lungs (Figure 2) [1].

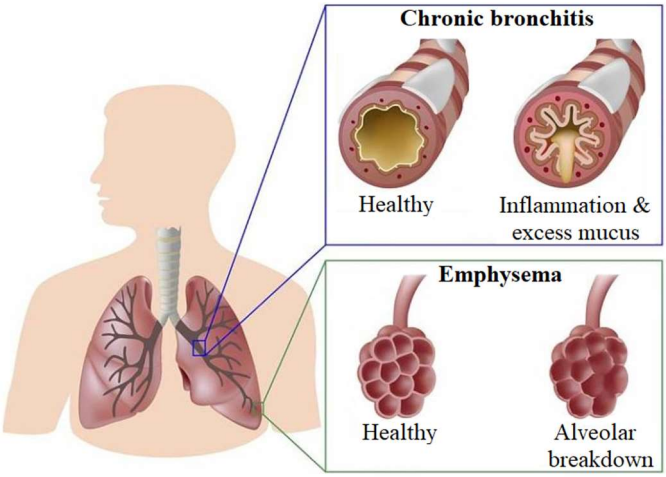

Figure 2. Airway and alveolar abnormalities in COPD.

The repeated injury and repair of the air wall itself may also lead to an excessive production of muscle and fibrous tissue [4]. This restricts the airflow to the alveoli and diminishes the ability of the small airways to remain open during exhalation. The latter results in trapping of air with a low partial pressure of oxygen and a high partial pressure of carbon dioxide. The trapped air does not contribute to gas exchange in the alveoli, while deoxygenated blood still perfuses through the pulmonary capillaries.

Chronic inflammation also leads to a loss of alveolar attachments and a destruction of the alveoli [3]. This is commonly referred to as emphysema and causes adjacent alveoli to merge and form larger air sacs (Figure 2). This greatly limits the surface area involved in gas exchange and the diffusion of oxygen, so the individual becomes short of breath [2]. Noxious exposure also causes a destruction of ciliary function. This leads to chronic cough and increased sputum production as the lungs can no longer remove mucus with associated debris or pathogens. This causes further narrowing, inflammation, and irritation of the airways. Dyspnoea (shortness of breath), chronic cough, and increased sputum production are therefore the most characteristic symptoms of COPD [3].

2.2.1 THE BURDEN OF COPD

COPD is a complex and heterogeneous disease that is responsible for substantial and increasing morbidity, mortality and healthcare expenditures worldwide [5]. It has been estimated that 384 million people have COPD, with a global prevalence of 11.7% [6]. COPD has climbed from the fourth leading cause of death in 1990 to the third

leading cause of death in 2010 [7]. Each year, COPD is responsible for 3 million deaths, representing nearly 6% of all deaths globally [8]. The burden of the disease is projected to increase further over the coming decades due to continued exposure to COPD risk factors in developing countries and ageing populations in high-income countries [9]. It is estimated that as many as 4.5 million deaths from COPD will occur in 2030 [10].

The direct cost of respiratory diseases in the European Union is €55 billion, where COPD accounts for the majority of the costs at €23.3 billion. If the indirect costs and monetized value of disability-adjusted life years are included, the total cost of COPD is estimated at €141 billion, equal to 37% of the total costs of respiratory diseases [11]. The global burden of the disease is primarily attributable to COPD exacerbations, and there is a striking relationship between disease progression and healthcare cost [3].

2.2.2 RISK FACTORS FOR COPD

COPD develops from a complex interaction between environmental exposure and genetics [3]. Cigarette smoke is the most common risk factor for COPD, and smokers are associated with a higher prevalence of respiratory symptoms, lung function abnormalities, a greater annual rate of decline in lung function, and a greater COPD mortality rate compared to non-smokers [12,13]. Other types of tobacco use, such as pipes, cigars, and water pipes, and marijuana use are also risk factors for COPD [14–16]. Passive exposure to cigarette smoke may also contribute to respiratory symptoms and COPD [17]. Occupational exposure to organic and inorganic dusts, chemical agents, and fumes are also often over-looked risk factors for COPD even in non-smokers [18–20].

Age is also often presented as a risk factor for COPD, as COPD predominantly develops after the age of 40 [21]. However, it is not clear whether age reflects the sum of exposure throughout life or whether ageing of the lungs appears similar to the structural changes associated with COPD [22]. Previous studies have often reported a greater COPD prevalence and mortality among men, but recent data from developed countries suggest that the prevalence of COPD is now almost equal between men and women [23]. Some studies have even suggested that the effects of tobacco smoke lead to a more severe disease in women compared to men for an equal quantity of smoked cigarettes [24–26].

Asthma might also be a risk factor for developing COPD, as adults with asthma have been found to have a 12-fold higher risk of developing COPD over time compared to adults without asthma after adjusting for smoking [27]. Another study reported that

approximately 20% of subjects with childhood asthma later developed persistent airflow limitation and decreased diffusion capacity [28]. Lung growth and development as well as respiratory infections in early childhood also have the potential to increase the risk of COPD developing later in life [29–31]. A severe hereditary alpha-1-antitrypsin deficiency is also among the best documented genetic risk factors for COPD [3,32]. Furthermore, there appears to be a significant familial risk of airflow limitation that has been observed in smokers who have siblings with severe COPD [33].

2.2.3 COPD DIAGNOSIS AND ASSESSMENT

COPD should be considered in any patient with shortness of breath, chronic cough, and sputum production with a history of risk factors. In a clinical context, a post-bronchodilator spirometry is required to diagnose COPD, where a ratio between the forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) of less than 70% confirms the presence of persistent airflow limitation [3]. A reduction in the FEV_1/FVC ratio is correlated with the extent of inflammation, mucus, and fibrosis of the airways [34]. Thus, COPD is diagnosed in a patient with persistent airflow limitation who presents with appropriate symptoms with significant exposure to noxious stimuli [3].

Assessment of COPD involves determining the level of airflow limitation, the impact on health status, and the risk of future events (exacerbations, hospital admissions or death) to guide therapy [3]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification can be used to categorize the level of airflow limitation, from 1 (mild) to 4 (very severe), based on the predicted FEV_1 percentage compared to an average healthy person of the same age, race, height, and gender [3]. The most widely used measures for assessing symptoms are the Modified British Medical Research Council (mMRC) dyspnoea scale, from 0 to 4, for assessing breathlessness [35] and the 8-item COPD Assessment Test (CAT) to measure health status impairment in COPD using a score from 0 to 40 [36]. It is also important to determine the history of exacerbations to assess the risk of future exacerbations [3,37].

2.2.4 COPD EXACERBATIONS

COPD is characterized by exacerbations, which are defined as an acute worsening of respiratory symptoms that results in additional therapy [3,38,39]. Patients with COPD with moderate airflow limitation may experience exacerbations that require treatment with antibiotics and/or systemic corticosteroids, but the risk of exacerbations is significantly higher for severe and very severe patients [3]. Exacerbations are important in managing COPD, as they have a negative impact on health status, rate of

hospitalization and readmission, and disease progression [38,40,41]. Exacerbations are complex and usually associated with an increase in airway inflammation, mucus production, and gas trapping. These changes cause increased dyspnoea, which is the key symptom of an exacerbation. Exacerbations are classified as mild (treated with short-acting bronchodilators), moderate (treated with short-acting bronchodilators plus antibiotics and/or oral corticosteroids), and severe (requiring hospitalization or an emergency room visit). Severe exacerbations may result in acute respiratory failure [3].

Exacerbations are mainly initiated by respiratory viral infections, while bacterial infections and environmental exposure (e.g., pollutants or temperature) may also initiate and/or amplify such events [42]. Lower respiratory tract viral infections may directly damage the airways in patients with COPD, resulting in a loss of the ciliated epithelium with increased inflammation and mucus production [43]. In addition to a further narrowing of the airways, the alterations may also promote a secondary bacterial infection, causing more inflammation and damage [44]. Exacerbations associated with viral infections, e.g., the human rhinovirus as commonly seen during winter [42], are often more severe, longer lasting, and cause more hospitalizations [3]. Symptoms during an exacerbation usually lasts 7 to 10 days, with systemic inflammation reaching baseline levels after 14 days, although severe exacerbations may last longer [45,46].

2.2.5 MANAGEMENT OF COPD

Once COPD is diagnosed, effective management should aim towards improving the patient's health status while reducing the current symptoms and the future risk of exacerbations. The identification and reduction of noxious exposure are important for treating and preventing the development of COPD [3]. Cigarette smoking is the most prevalent risk factor, and smoking cessation should be highly encouraged for all patients with COPD who continue to smoke [47]. Influenza vaccination can reduce lower respiratory tract infections [48] and mortality in patients with COPD [49]. This vaccination is especially important in elderly patients [50]. The pneumococcal vaccine can also provide significant protection against community-acquired pneumonia and reduce the likelihood of a COPD exacerbation [51]. Pharmacological treatment of COPD can reduce symptoms and the frequency of exacerbations while increasing the health-related quality of life and exercise tolerance for patients with COPD [52]. The treatment primarily involves the use of bronchodilators (relaxes airway smooth muscle to open up the airways), antibiotics (reduce exacerbation rate), and inhaled corticosteroids (reduce airway inflammation) [3]. In very severe cases of COPD with chronic respiratory failure, oxygen therapy has been shown to increase survival [53].

Additionally, COPD often coexists with other diseases (comorbidities), and it has been increasingly recognized that COPD is part of a systemic disease [54,55]. Comorbidities of COPD include cardiovascular diseases [56], lung cancer [57], diabetes [58], osteoporosis [59], loss of skeletal muscle [60], renal failure [61], anxiety and depression [62], among others [63,64], that all have negative effects on the disease prognosis. Comorbidities may arise from sharing risk factors with COPD, e.g., smoking, while some diseases may increase the risk of others [3]. Comorbidities of COPD have also been shown to affect exacerbations [65]. For instance, heart failure is frequent and often unrecognized in patients hospitalized for COPD exacerbations [66]. COPD is therefore increasingly recognized as the pulmonary component of multimorbidity, which should be taken into consideration when managing COPD [3].

COPD exacerbations are especially important to treat, with the goal of minimizing the negative impact of the current exacerbation and preventing subsequent exacerbations [3]. More than 80% of exacerbations may be managed outside the hospital with bronchodilators, corticosteroids, and antibiotics [45,67,68]. Exacerbations with severe symptoms of acute respiratory failure also require ventilator support and hospitalization [3]. The in-hospital mortality for patients admitted for COPD exacerbations with hypercapnic respiratory failure has been reported to be 11%, with a 33% mortality after 180 days and a 49% mortality after 2 years [69]. Early treatment has been associated with faster recovery, and patients with a high proportion of reported exacerbations have better health-related quality of life and a decreased risk of emergency hospitalization [70]. However, patients with COPD generally have a poor understanding of their disease and symptoms, thus delaying presentation or failing to seek treatment during exacerbation [40]. It has been estimated that approximately 40% to 70% of exacerbations are not reported to healthcare professionals [40,70,71]. This is of great concern, as untreated exacerbations are associated with a significant worsening of respiratory symptoms and health-related quality of life, a longer recovery time from the exacerbation and a faster deterioration in lung function [72,73]. Therefore, there is an urgent demand for developing tools that emphasize the early identification of COPD exacerbations [74]. Telemonitoring of patients with COPD and the application of predictive analytics have been proposed as a potential solution to predict exacerbations prior to onset.

2.3 TELEMONITORING FOR MANAGING COPD

Telemonitoring is an application of telecommunication technology that allows communication and remotely collected clinical data to be exchanged between the patient's home and the healthcare provider on a regular basis. Telemonitored data can involve physiological measurements (e.g., spirometry, blood pressure, oxygen

saturation) and/or respiratory symptom questionnaires (e.g., ‘more cough than usual’, ‘sputum purulence’, etc.). Telemonitoring has the potential to improve the quality of life in patients with COPD while reducing exacerbations and hospital admissions through personalized counselling and feedback from the healthcare provider [75]. Although telemonitoring may have a positive effect on empowerment and quality of life among patients with COPD, the effect on exacerbations and hospitalizations is unconvincing [76–82]. However, this emerging data-rich environment encourages the use of predictive analytics to assist the management of COPD though predicting exacerbations prior to onset [83]. This could facilitate an early and preventive treatment of exacerbations and presumably lead to better patient outcomes and a decreased risk of hospitalization [70].

2.3.1 PREDICTIVE ANALYTICS IN CONCEPT

Within predictive analytics, statistical or machine learning methods are used to model an outcome based on a set of informative predictors [84]. A fictional example of a binary outcome could be to predict whether a patient with COPD will encounter an exacerbation in the upcoming week. For this prediction example, the model could be based on measurements of blood pressure and oxygen saturation from the past week. The model would rely on patterns in the data set to make decisions or predictions of the outcome [84]. Generally, the development of a predictive model starts with acquiring a representative data set including a desired outcome and a series of variables. Pre-processing follows to condition the data set with the extraction of potential predictors. Subsequently, the task is to identify the subset of predictors that is most informative of the outcome and then select an appropriate classification algorithm to base the model upon. One can think of a classification algorithm as a function approximation problem that maps an input of predictors to the corresponding outcome. Finally, the model should be tested on previously unseen data, where the model is only given the predictors as input. The true performance of the model is then determined by comparing the model-predicted outcomes to the true outcomes [85].

For a binary outcome, there are four possibilities when comparing the model-predicted outcomes to the true outcomes (Table 1). There are two correct cases in which the model correctly predicts either the positive outcome (e.g., COPD exacerbation in the upcoming week encoded as a ‘1’) or the negative outcome (no COPD exacerbation encoded as a ‘0’). These two cases are referred to as *true positive* and *true negative*, respectively. If the model-predicted outcome is not in accordance with the true outcome, there can either be a *false positive* (wrongly predicted COPD exacerbation) or a *false negative* (wrongly predicted absence of COPD exacerbation). The performance of a model is often expressed using several statistical measures,

often referred to as classification rates. These include sensitivity, specificity, positive predictive value, negative predictive value, and accuracy, which are all extracted based on the ratios among true positives, true negatives, false positives, and false negatives (Table 1) [86].

		True outcome		
		1	0	
Model-estimated outcome	1	True positive (TP)	False positive (FP)	Positive predictive value: $\frac{TP}{TP + FP}$
	0	False negative (FN)	True negative (TN)	Negative predictive value: $\frac{TN}{TN + FN}$
		Sensitivity: $\frac{TP}{TP + FN}$	Specificity: $\frac{TN}{TN + FP}$	Accuracy: $\frac{TP + TN}{TP + FP + FN + TN}$

Table 1. Prediction outcome and classification rates of model performance. Inspired from [86].

The classification rates presented above are considered threshold-specific and are therefore often a poor metric of the general performance of a model. A model will often output the probability of the positive outcome (COPD exacerbations in the upcoming week) instead of the outcome itself. A threshold will then determine the probability at or above which the positive class will be predicted based on the model. A low threshold would, for example, often result in a model with high sensitivity and low specificity due to many true positives at the cost of many false positives [86].

The receiver operating characteristic (ROC) curve is a technique for visualizing the general performance of a model (Figure 3). A ROC curve is generated by varying the

probability threshold from 0 to 1 while displaying the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) [86].

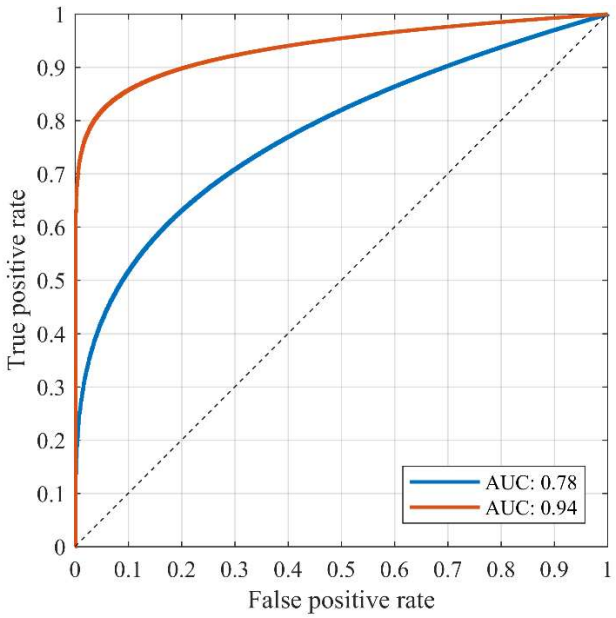

Figure 3. Examples of receiver operating characteristic curves for two models and their corresponding area under the curve values.

A common method to quantify the general performance of a model is the area under the ROC curve (AUC). The AUC ranges from 0 to 1, where an AUC of 1 corresponds to a perfect classification and an AUC of 0.5 (the diagonal line, Figure 3) is what can be expected from random guessing [86,87]. As exemplified in Figure 3, the red model is able to maintain a high specificity while correctly classifying a large portion of the positive outcomes compared to the blue model. If a specificity of 0.90 is acceptable, the red model yields a sensitivity of 0.86 as opposed to 0.52 for the blue model.

2.3.2 PREDICTIVE MODELS FOR COPD EXACERBATIONS

Several previous studies have attempted to develop models for predicting upcoming COPD exacerbations to facilitate an early and preventive treatment [88–94]. These studies are based on telemonitored data from physiological measurements and/or respiratory symptom questionnaires (Table 2).

Study	**Predictor(s)**	**Classification algorithm**	**Performance**
Jensen 2012 [88]	- Oxygen saturation	Linear discriminant classification	Sens: 0.70 Spec: 0.95 NPV: 0.88 PPV: 0.86 AUC: 0.73
van der Heijden 2013 [89]	- Lung function - Oxygen saturation - Respiratory symptoms	Bayesian network	Sens: 0.88 Spec: 0.80 Acc: 0.81 AUC: 0.87
Fernandez-Granero 2014 [90]	- Respiratory symptoms	Probabilistic neural network	Sens: 0.81 Spec: 0.94 Acc: 0.88 NPV: 0.86 PPV: 0.92
Mohktar 2015 [91]	- Blood pressure - Body temperature - Bodyweight - Heart rate - Lung function - Respiratory symptoms	Classification and regression tree (CART)	Sens: 0.61 Spec: 0.80 Acc: 0.72
Sanchez-Morillo 2015 [92]	- Respiratory symptoms	K-means clustering	Sens: 0.75 Spec: 0.90 Acc: 0.85 NPV: 0.88 PPV: 0.79 AUC: 0.84
Riis 2016 [93]	- Blood pressure - Heart rate - Oxygen saturation	K-nearest neighbour	Sens: 0.73 Spec: 0.74 Acc: 0.74 NPV: 0.78 PPV: 0.69
Shah 2017 [94]	- Heart rate - Oxygen saturation - Respiratory rate	Logistic regression	Sens: 0.60-0.80 Spec: 0.68-0.36

Table 2. Previously reported predictive models for COPD exacerbations based on data collected through telemonitoring. Abbreviations: Sens: sensitivity, Spec: specificity, Acc: accuracy, NPV: negative predictive value, PPV: positive predictive value, AUC: area under the curve.

Maintaining a high specificity is crucial, as the annual rate of exacerbations has been reported to be 2.4 to 2.7 exacerbations per patient per year in patients with moderate to severe COPD [40,45]. This results in a natural imbalance between outcomes when predicting an upcoming exacerbation. The studies by Jensen et al. [88], Fernandez-Granero et al. [90], and Sanchez-Morillo et al. [92] all reported good performances with specificities in the range of 0.90 to 0.95 with corresponding sensitivities in the range of 0.70 to 0.81 using a linear discriminant classification algorithm, a probabilistic neural network, and a k-means clustering algorithm, respectively.

However, as also concluded in a recent systematic review, predictive models have yet to achieve clinically useful levels of sensitivity, specificity, and accuracy [83]. Furthermore, it is challenging to compare the results of previously reported predictive models as different methods used for assessing the performance can highly influence the generalizability of the results. This calls for a new approach in developing predictive models before they can be successfully implemented in telemonitoring of patients with COPD.

2.3.3 UNEXPLORED POTENTIALS

Recently reported evidence has indicated that previous exacerbations are associated with an increased risk of future exacerbations [95,96]. This potential has yet to be investigated, as no previous studies, to our knowledge, have included information of previous exacerbations as a predictor. It is also well established that COPD exhibits seasonal variation and that environmental factors increase the risk of COPD exacerbation [97–100]. One might therefore further hypothesize that the frequency of COPD exacerbations among patients in general is potentially associated with an increased risk of exacerbation for the individual patient. This has also yet to be included in the prediction of exacerbations in patients with COPD.

CHAPTER 3. THESIS OBJECTIVE

This chapter presents the overall objective of the thesis and the specific objectives addressed in the four thesis studies.

COPD is a major socio-economic burden that is associated with significant morbidity and mortality worldwide. The impacts of COPD are mainly attributable to exacerbations, which are important to treat early and prevent subsequently. A substantial portion of unreported exacerbations is a challenge in the management of COPD. Telemonitoring and the application of predictive analytics have been suggested as a potential solution, but predictive models have yet to obtain a level of clinical relevance. There is unexplored potential in including information on previous exacerbations, both within and between patients with COPD, which could potentially improve the prediction of COPD exacerbations. This leads to the following objective:

The objective of this thesis was to improve methods for assessing predictive models and to explore the potential of including information on previous exacerbations in the prediction of upcoming exacerbations in patients with COPD.

Two data sets were obtained to form the basis for developing predictive models to address the thesis objective (Table 3). Data set A contained a large cohort of telemonitored patients with COPD from the TeleCare North Trial [101]. Data set B contained a small cohort with a higher frequency of measurements from a telemedicine project at the Department of Pulmonary Medicine, Aalborg University Hospital [102].

Data set	Patients included	Variables	Frequency of measurements	Exa/ctrl periods
A	57	- Blood pressure - Heart rate - Oxygen saturation - Weight - Respiratory symptoms	1-2 per week	84/1652
B	9	- Blood pressure - Heart rate - Oxygen saturation	≥3 per week	17/398

Table 3. Summary of the processed data sets used for developing predictive models for addressing the thesis objective. Exa/ctrl (exacerbation/control) periods are periods with or without a subsequent occurrence of exacerbation.

3.1 PAPER-SPECIFIC OBJECTIVES

The thesis objective was addressed in four separate studies that were presented in four papers. The specific aims of the studies are presented in Table 4.

Study	Objective	Data set
I	To explore whether the patient and population incidence of exacerbation in a large cohort of patients with COPD contain predictive information of upcoming exacerbations. Paper I: *Population exacerbation incidence contains predictive information of acute exacerbations in patients with chronic obstructive pulmonary disease in telecare*	A
II	To propose a proper validation method for developing predictive models, as overfitting is a prevalent problem in small and imbalanced data sets. Paper II: *Double cross-validation with feature selection: An overfit-resistant and transparent validation method in a small imbalanced data set for predicting exacerbations in patients with COPD*	B
III	To provide a systematic comparison of classification algorithms for predicting COPD exacerbations in order to assess the sensitivity of the choice of classification algorithm. Paper III: *A systematic approach to sensitivity-analysis of choice of classification algorithm for prediction of COPD exacerbations*	B
IV	To explore whether the inclusion of previous model-estimated probabilities of exacerbation can improve the prediction of upcoming COPD exacerbations (as a proxy for the patient and population incidence of exacerbation). Paper IV: *A two-layer probabilistic model shows promising results for predicting COPD exacerbations for patients in telemonitoring*	B

Table 4. The four studies and their specific objectives to address the thesis objective.

CHAPTER 4. SUMMARY OF PAPERS

This chapter provides a summary of each of the four papers that have been included in the thesis.

4.1 PAPER I

Title: Population exacerbation incidence contains predictive information of acute exacerbations in patients with chronic obstructive pulmonary disease in telecare

Introduction: No previous studies have reported predictive models that include information on previous exacerbations in the prediction of upcoming exacerbations. The objective of this study was to investigate whether the patient and population incidence of exacerbation contain predictive information of acute COPD exacerbations in combination with physiological measurements and symptom questionnaires.

Methods: Weekly submissions of blood pressure, oxygen saturation, heart rate, weight, and a respiratory symptom questionnaire were obtained from 578 patients with COPD, who constituted the intervention group out of 1225 patients who were included in the TeleCare North Trial. The data were structured into 30-day periods and labelled either as an exacerbation period if the patient was hospitalized shortly after the period or a control period if no hospitalization occurred. Periods required at least three measurements during the 30 days and were otherwise excluded. Twenty-nine predictors were extracted for every period and involved mathematical extractions of the physiological parameters, a symptom-defined exacerbation based on the questionnaire, patient incidence of exacerbation in the last three, six, and twelve months, and population incidence of exacerbation within the last month. Two predictive models were then developed based on logistic regression to differentiate between exacerbation periods and control periods. One model was provided with potential predictors from the patient and population incidence of exacerbation, while the other model was not. An exhaustive predictor search was used in each of the models to identify up to four of the most informative predictors based on five-fold cross-validation. Each model was evaluated by AUC using two-fold cross-validation, and the models were compared to assess whether the inclusion of the patient or population incidence of exacerbation improved the prediction of acute COPD exacerbations.

Results: A total of 84 exacerbation periods and 1652 control periods among 57 patients were extracted from the data set. The model without the patient and population incidence of exacerbation was based on oxygen saturation, weight, and symptom-defined exacerbation and yielded a mean AUC of 0.63. In comparison, the model with the patient and population incidence of exacerbation as potential predictors was based on oxygen saturation, weight, and population incidence of exacerbation and resulted in a mean AUC of 0.74. The difference in AUC was statistically significant ($p < 0.05$), and the latter model obtained a sensitivity of 0.67 at a specificity of 0.70.

Conclusion: Population exacerbation incidence was considered an informative predictor for acute COPD exacerbations. The population exacerbation incidence, in combination with physiological parameters, increased the performance of a predictive model compared with a conventional model based on only physiological parameters and respiratory symptom questionnaires.

4.2 PAPER II

Title: Double cross-validation with feature selection: An overfit-resistant and transparent validation method in a small imbalanced data set for predicting exacerbations in patients with COPD

Introduction: Over-fitting is a prevalent problem when validating the performance of predictive models, especially in small imbalanced data sets. The objective of this study was to compare models based on double cross-validation with feature selection to models based on conventional cross-validation and simple validation (over-fitting).

Methods: Daily to weekly measurements of blood pressure, oxygen saturation, and heart rate were obtained from 108 patients with COPD included in a telemedicine study at the Department of Pulmonary Medicine, Aalborg University Hospital. The measurements were structured into 14-day periods with a minimum of three measurements in both the first and the second week. If an exacerbation, defined as a patient-initiated self-medication plan or a hospitalization, occurred within three days of the last measurement, the period was defined as an exacerbation period and otherwise as a control period. Patients with no exacerbation periods were excluded. Seventy potential predictors were obtained based on mathematical extractions of measurements. Predictive models for COPD exacerbations were allowed up to three predictors and were developed based on four different classification algorithms: logistic regression, naive Bayes, random forest, and support vector machine. For each of the four classification algorithms, the models were validated based on simple validation, cross-validation, and double cross-validation with feature selection. The three validation methods were compared by ROC curves and AUC.

Results: Nine patients remained after structuring the data set, resulting in a total of 17 exacerbation periods and 398 control periods. By visual inspection of the ROC curves, the performance of the models based on cross-validation were between that of the models based on simple validation and double cross-validation with feature selection for all four classification algorithms. This was also reflected in the AUC values (Table 5).

Classification algorithm	**SV**	**CV**	**DCVFS**
Logistic regression	0.89	0.86	0.81
Naive Bayes	0.91	0.85	0.72
Support vector machine	0.89	0.85	0.72
Random forest	1.00	0.88	0.78

Table 5. AUC values for the models based on simple validation (SV), cross-validation (CV), and double cross-validation with feature selection (DCVFS) among the four classification algorithms.

Conclusion: Double cross-validation with feature selection was considered a more robust validation method than conventional cross-validation when predicting COPD exacerbations in a small imbalanced data set.

4.3 PAPER III

Title: A systematic approach to sensitivity-analysis of choice of classification algorithm for prediction of COPD exacerbations

Introduction: A comparison of predictive models reported in previous studies is problematic due to varying data sets, different validation methods, and different classification algorithms being applied. No previous studies have reported a systematic comparison of classification algorithms using the same data set and the same validation method. The objective of this study was to systematically compare classification algorithms to assess the sensitivity to the choice of classification algorithm for predictive models for COPD exacerbations.

Methods: Data from 108 patients with COPD were obtained from a telemedicine study at the Department of Pulmonary Medicine, Aalborg University Hospital. Patients continuously submitted home measurements of blood pressure, oxygen saturation, and heart rate, which were retrospectively structured into 14-day periods. Periods required at least three measurements in both the first and second week. Exacerbations were defined as a patient-initiated self-medication plan or a hospitalization. The periods further required no more than three days between the last measurement in the period and the occurrence of an exacerbation. Control periods were defined as periods with no subsequent exacerbation. Patients with no exacerbation periods were excluded from the study. Mathematical extractions of measurements resulted in a total of seventy-five potential predictors. Nine classification algorithms were included in the sensitivity analysis: linear discriminant classification, logistic regression, classification tree, random forest, support vector machine with both a linear kernel and a radial basis function kernel, naive Bayes, and k-nearest neighbour with both three and five nearest neighbours. A predictive model was developed for each classification algorithm with three predictors based on double cross-validation with feature selection. The predictive models were evaluated and compared by AUC and by sensitivity at a fixed specificity of 0.95.

Results: A total of nine patients with 17 exacerbation periods and 398 control periods remained after data structuring. The comparison of the predictive models revealed that the different classification algorithms resulted in AUC values ranging from 0.52 to 0.81 with a mean value of 0.71±0.10. The sensitivity at a fixed specificity of 0.95 ranged from 0.18 to 0.65 with a mean value of 0.37±0.18.

Conclusion: Prediction models for COPD exacerbations are highly sensitive to the choice of classification algorithm. Several classification algorithms should therefore be compared systematically when developing predictive models.

4.4 PAPER IV

Title: A two-layer probabilistic model shows promising results for predicting COPD exacerbations for patients in telemonitoring

Introduction: Population exacerbation incidence was previously shown to be a predictor of upcoming exacerbations in patients with COPD. However, this study was limited by the frequency of measurements and by that it is rarely possible to obtain continuous access to medical records to determine the day to day population exacerbation incidence. As an alternative to accessing medical records, this study attempted to use previous model-estimated probabilities of exacerbations as a proxy for the patient and population incidence of exacerbation. The objective of this study was to investigate whether a two-layer probabilistic model can increase classification rates compared to a conventional one-layer model for predicting COPD exacerbations.

Methods: Continuous measurements of blood pressure, oxygen, saturation, and heart rate were obtained from a telemedicine study performed at the Department of Pulmonary Medicine, Aalborg University Hospital, which included 108 patients with COPD. The measurements were retrospectively structured into 14-day periods requiring at least three measurements in each of the weeks in the period. Exacerbation periods were defined as a patient-initiated self-medication plan or a hospitalization occurring within three days after the last measurement in the period. Control periods were defined as periods with no subsequently occurring exacerbations. Patients with no exacerbation periods were excluded from the study. Taking inspiration from previous work, nine classification algorithms were included in a comparison between a two-layer probabilistic model and a conventional one-layer model. The classification algorithms included were linear discriminant classification, logistic regression, classification tree, random forest, support vector machine with both a linear kernel and a radial basis function kernel, naive Bayes, and k-nearest neighbour with both three and five nearest neighbours. The one-layer models were based on three out of seventy-five potential predictors from the mathematical extractions of measurements and were developed based on double cross-validation with feature selection. The estimated probabilities of exacerbation from each one-layer model were used to extract eleven potential predictors in the two-layer model, which also allowed three predictors and was developed using double cross-validation with feature selection. The two models were compared by AUC and sensitivity at a fixed specificity of 0.95 for each classification algorithm.

Results: Nine patients remained after structuring the data set, resulting in a total of 17 exacerbation periods and 398 control periods. For all classification algorithms, the two-layer model increased the AUC value and by a mean of 0.11. The sensitivity was

also increased in four out of nine cases, decreased in two cases, and was unchanged in the remaining three cases. However, the sensitivity was increased overall by a mean of 0.13. The best one-layer model exhibited a sensitivity of 0.65 at a specificity of 0.95 based on the linear discriminant classification algorithm, and the best two-layer model yielded a sensitivity of 0.94 at a specificity of 0.95 based on the support vector machine algorithm with a radial basis function kernel.

Conclusion: A two-layer probabilistic model can significantly increase the classification rates to a level of clinical relevance compared to a conventional one-layer model for predicting COPD exacerbations.

CHAPTER 5. THESIS PAPERS

This chapter presents the four papers that have been included in the thesis. The full papers are included in the appendix of the full version of the thesis.

5.1 PAPER I

Population exacerbation incidence contains predictive information of acute exacerbations in patients with chronic obstructive pulmonary disease in telecare

Thomas Kronborg, Lasse Mark, Simon L. Cichosz, Pernille Heyckendorff Secher, Ole Hejlesen

Published in: International Journal of Medical Informatics (2019)

5.2 PAPER II

Double cross-validation with feature selection: An overfit-resistant and transparent validation method in a small imbalanced data set for predicting exacerbations in patients with COPD

Thomas Kronborg, Simon L. Cichosz, Ole Hejlesen

Submitted to: Computers in Biology and Medicine (2019)

5.3 PAPER III

A systematic approach to sensitivity-analysis of choice of classification algorithm for prediction of COPD exacerbations

Thomas Kronborg, Simon L. Cichosz, Ole Hejlesen

Submitted to: Computers in Biology and Medicine (2019)

Computers in Biology and Medicine

5.4 PAPER IV

A two-layer probabilistic model shows promising results for predicting COPD exacerbations for patients in telemonitoring

Thomas Kronborg, Simon L. Cichosz, Ole Hejlesen

Submitted to: International Journal of Medical Informatics (2019)

CHAPTER 6. DISCUSSION

This chapter will initially summarize the main findings and discuss the methodological considerations of the four papers included in the thesis. It will then focus on the significant insights that the thesis has provided and the future perspectives.

6.1 SUMMARY OF THE MAIN FINDINGS

The objective of this PhD study was to improve methods for assessing predictive models and to explore the potential of including information on previous exacerbations in the prediction of upcoming exacerbations in patients with COPD. Paper I presented how the population exacerbation incidence is an informative predictor for COPD exacerbations in combination with physiological parameters. Paper II encouraged the use of double cross-validation with feature selection, as this validation method was considered more robust than conventional cross-validation for predicting exacerbations in a small imbalanced data set. Paper III showed that prediction models for COPD exacerbations are highly sensitive to the choice of classification algorithm and that several classification algorithms should therefore be compared when developing predictive models. Papers IV demonstrated how previous model-estimated probabilities of exacerbation can be implemented in a two-layer probabilistic model to significantly increase classification rates for COPD exacerbations compared to a conventional model.

6.2 METHODOLOGICAL CONSIDERATIONS

The strengths of the study presented in Paper I were a relatively large sample of patients with COPD and that exacerbations were defined from hospitalizations. The sample size (n=57) increased the generalizability of the results, and a hospitalization-defined exacerbation ensured that the models were focused on predicting only severe exacerbations, which cannot be managed outside of hospital [3]. These cases are also associated with the largest negative impact on disease progression and substantial healthcare costs [103]. An additional strength of the study was that two predictors were shared among the models with and without patient and population exacerbation incidence as potential predictors. This supports the fact that population exacerbation incidence is an informative predictor of COPD exacerbations and led to a significant increase in AUC. The weaknesses of the study presented in Paper I primarily included a low frequency of measurements and a time span of up to a week between the last measurement and the occurrence of exacerbation. This resulted in data structuring,

where periods only required three measurements in a 30-day period, while some patients measured two times a week. This presumably induced a large variance within the predictors, as some periods were based on three measurements, while other periods were based on up to eight or nine. The time span of up to a week could also have complicated the prediction, as the most informative predictors for predicting exacerbations may differ between one day and seven days prior to onset. These weaknesses could have caused the low performance of the models compared to other studies that used a higher frequency of measurements. Ideally, the frequency of measurements would have been increased and standardized, which would have provided better conditions for developing a prediction model.

Common strengths of the studies presented in Papers II-IV included a high frequency of measurements and no more than three days between the last measurement and the occurrence of an exacerbation. This allowed a minimum of six measurements within a 14-day period and therefore more information closer to the occurrence of the exacerbation, as opposed to the study presented in Paper I. Similarly, a common weakness of the studies was the limited sample size (n=9), which potentially yielded results with a low generalizability and prevented the inclusion of more sophisticated algorithms that require large training sets. To address this weakness, models were developed and validated using double cross-validation with feature selection, and the results were based on several classification algorithms, as proposed in Papers II and III.

The additional strength of the study presented in Paper II was that the results showed the same tendency with all four models based on different classification algorithms. The primary weakness of this study was that there was no independent test set. An independent test set could have provided insights into the degree of generalizability, where potential over-fitting of the different validation methods would have been reflected in a lower performance on the independent test set.

Additional strengths of the study presented in Paper III were that several classification algorithms were systematically compared using the same data set and the same validation method. The classification algorithms were the only variable, which resulted in a fair comparison as opposed to comparing models from previous studies with different data sets, validation methods, and classification algorithms. A weakness of this study was, presumably, that not all classification algorithms were developed to their full potential, as only cost-sensitive learning was used. Several of the classification algorithms can be fine-tuned in many ways, but this was considered too comprehensive to include in the study.

The strength of the study presented in Paper IV was that the probabilistic two-layer model increased the classification rates across all the classification algorithms by a substantial margin. A weakness of the study was the potential of selecting different predictors when developing models based on double cross-validation with feature selection. This prevented the possibility of providing a general predictive model, which could be implemented in the telemonitoring of patients with COPD. However, the sample size was not considered sufficient for providing a general model.

In the presented studies, exacerbations were defined from either a patient-initiated self-medication plan and/or a hospitalization. Prior to severe exacerbation, the predictors presumably vary more from the patient's stable state compared to mild or moderate exacerbations, but this is uncertain. Ideally, exacerbations should have been stratified by severity, according to the required treatment [3], and further by length of stay or healthcare costs for severe exacerbations. It remains uncertain whether defining exacerbations from a patient-initiated self-medication plan is accurate, as patients with COPD have been reported to have a poor understanding of their disease and symptoms [40]. Patients with a good understanding of their disease and symptoms may be able to reduce the impact of a potential moderate or severe exacerbation through early recognition. In contrast, an unreported, and therefore untreated, mild or moderate exacerbation might have a substantial impact in patients with a poor understanding [104]. In this case, a two-layer probabilistic model may be a better alternative for defining exacerbations based on the variations from the patient's stable state. This encourages research towards a model-based data-driven definition for mild and moderate exacerbations in combination with the hospitalization-based definition for severe exacerbations.

6.3 SIGNIFICANCE OF RESEARCH

It is challenging to compare the results of the studies included in this thesis to those of previously reported models for COPD exacerbations due to varying data sets and class imbalances, the use of different classification algorithms and different validation methods, and variations in the window for predicting an exacerbation. However, Shah et al. [94] reported a predictive model with a mediocre performance, a sensitivity of 0.60 at a specificity of 0.68, based on logistic regression. The studies reported by Riis et al. [93] and Mohktar et al. [91] resulted in models having slightly better performances, with a sensitivity of 0.73 at a specificity of 0.74 based on the k-nearest neighbour algorithm and a sensitivity of 0.61 at a specificity of 0.80 based on the CART algorithm. Although the study by van der Heijden et al. [89] reported a good sensitivity of 0.88, based on a Bayesian network, a specificity of 0.80 would still cause a crippling amount of false positives in a telemonitoring setting given the natural

imbalance between periods with and without exacerbation. The models reported in the studies by Sanchez-Morillo et al. [92], Jensen et al. [88], and Fernandez-Granero et al. [90] all managed to obtain a moderate to a good sensitivity while maintaining a specificity of 0.90 or more. Sanchez-Morillo et al. [92] reported a sensitivity of 0.75 at a specificity of 0.90 based on the k-means clustering algorithm, and Jensen et al. [88] reported a sensitivity of 0.70 at a specificity of 0.95 based on the linear discriminant classification algorithm. The previously reported model with the best performance was described in the study by Fernandez-Granero et al. [90]. This model demonstrated a sensitivity of 0.81 at a specificity of 0.94 based on a probabilistic neural network.

Compared to previous work in the field, the model reported in Paper I resulted in a mediocre performance with a sensitivity of 0.67 at a specificity of 0.70 based on logistic regression. This was considered reasonable due to the discussed weaknesses of the study and as the study was focused on investigating whether the patient and population incidence of exacerbation contain predictive information for upcoming exacerbations. From the results reported in Paper IV, the best one-layer model exhibited a sensitivity of 0.65 at a specificity of 0.95 based on the linear discriminant classification algorithm, which was better or comparable to most previously reported models. The best two-layer model yielded a sensitivity of 0.94 at a specificity of 0.95 based on the support vector machine algorithm with a radial basis function kernel. The performance of this model was superior to that of previously reported models for predicting COPD exacerbations.

This thesis has presented results that have the potential to improve the state-of-the-art in the use of predictive analytics for managing COPD. Including information on previous exacerbations for predicting upcoming exacerbations resulted in a mean increase in the AUC of 0.12 for the prediction models presented in Paper I and a mean increase in the AUC of 0.11 for the prediction models presented in Paper IV. Comparing the two studies, a higher frequency of measurements generally resulted in predictive models with higher classification rates without including information on previous exacerbations. As presented in Paper I, the population exacerbation incidence was considered a good predictor in a large cohort with a lower frequency of measurements, while the patient exacerbation incidence was not included among the most informative predictors. In contrast, the model-estimated probability of exacerbation within patients was frequently selected as the most informative predictor in a smaller cohort with a higher frequency of measurements, while the model-estimated probability of exacerbation in the population was selected less, as shown in the study presented in Paper IV.

The results from the study presented in Paper IV indicated that a two-layer probabilistic model is the best approach for predicting COPD exacerbations. Moreover, the model based on the linear discriminant classification algorithm resulted in the highest classification rates. The first layer model should provide a model-estimated probability of exacerbation that is based on mathematical extractions from three or more weekly measurements of blood pressure, heart rate, and oxygen saturation from a two-week period. As a single peak in the model-estimated probability of exacerbation was not ideal, the second layer model should predict exacerbations based on the patient mean probability of exacerbation within the last week or two, the relative difference in probability of exacerbation from the patient median, and the population mean probability of exacerbation within the last week.

6.4 FUTURE PERSPECTIVE

Although the potential of including information on previous exacerbations in the prediction of upcoming exacerbations shows promising results, a general model for predicting COPD exacerbations has yet to be proposed. However, the results of this thesis have provided valuable insights into what should be included as predictors when developing a general predictive model. Proposing a general predictive model would require a large representative cohort with a high frequency of measurements. This might also reveal whether the model-estimated probability of exacerbation in the population could be an informative predictor in this case, similar to using the population exacerbation incidence presented in Paper I.

Based on the results included in the thesis, it has been agreed that the two-layer probabilistic model will be deployed into and tested in an already existing telemonitoring platform in the North Denmark Region. This has been chosen as a signature project supported by the Danish government, Local Government Denmark, and the Danish Regions as a part of a new national strategy to use artificial intelligence for improving healthcare [105]. A large cohort of patients with COPD will be instructed to perform frequent measurements of blood pressure, heart rate, and oxygen saturation. The study will be a cluster-randomized controlled trial in which healthcare personnel in the control group will manage patients with typical telemonitoring and those in the intervention group will manage patients assisted with a prediction model for exacerbations in combination with typical telemonitoring. The groups will be compared according to health-related quality of life, number of hospitalizations among the patients, and total healthcare costs to determine the true value of using predictive analytics for managing patients with COPD.

The two-layer probabilistic model approach will hopefully inspire future research in predictive analytics towards treating COPD as a multimodal systemic disease. Thus far, the prediction of COPD exacerbations has been approached by treating COPD solely as a respiratory disease. Using several first layer models for each of the comorbidities of COPD could potentially increase the generalizability of a prediction model to a much larger group of multimodal chronic patients. A good starting point would be to include a cardiovascular first layer model, as cardiovascular diseases are the most important comorbidities of COPD [56] and a frequent component in exacerbations [66].

CHAPTER 7. CONCLUSION

This chapter concludes the thesis.

In conclusion, double cross-validation should be used to develop and compare models based on several classification algorithms, and there is great potential in including information on previous exacerbations in the prediction of upcoming exacerbations in patients with COPD. Although this has been clinically recognized, no previous studies have included this aspect in the development of predictive models. This thesis has presented how information on previous exacerbations can be utilized to significantly increase classification rates for predictive models to a level of clinical relevance. This encourages the implementation of a predictive model in COPD telemonitoring to predict exacerbations and facilitate early and preventive treatment. This could potentially decrease the impact of exacerbations and reduce the morbidity, mortality, and socio-economic burden related to COPD.

LITERATURE LIST

[1] Ward JPT, Ward J, Leach RM. The respiratory system at a glance. 4th ed. New York: John Wiley & Sons Inc. 2015.

[2] Martini F, Nath JL, Bartholomew EF. Fundamentals of anatomy and physiology. 4th ed. San Francisco: Benjamin Cummings 2012.

[3] Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease 2019 report. Global Initiative for Chronic Obstructive Lung Disease, Inc.

[4] Rennard SI, Wachenfeldt Kv. Rationale and emerging approaches for targeting lung repair and regeneration in the treatment of chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society 2011;8(4):368-375. doi: 10.1513/pats.201102-019RM.

[5] Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. American Journal of Respiratory and Critical Care Medicine 2013;187(4):347-365. doi: 10.1164/rccm.201204-0596PP.

[6] Adeloye D, Chua S, Lee C, Basquill C, Papana A, Theodoratou E, et al. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. Journal of Global Health 2015;5(2)020415. doi: 10.7189/jogh.05-020415.

[7] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2095-2128. doi: 10.1016/S0140-6736(12)61728-0.

[8] Naghavi M, Wang H, Lozano R, Davis A, Liang X, Zhou M, et al. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;385(9963):117-171. doi: 10.1016/S0140-6736(14)61682-2.

[9] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Medicine 2006;3(11):e442.

[10] World Health Organization (WHO). Projections of mortality and causes of death, 2015 and 2030. Accessed November 2019. Available from: https:// www.who.int/healthinfo/global_burden_disease/projections2015_2030/en/.

[11] European Respiratory Society. European lung white book. Accessed November 2019. Available from: https://www.erswhitebook.org/.

[12] Kohansal R, Martinez-Camblor P, Agustí A, Sonia Buist A, Mannino DM, Soriano JB. The natural history of chronic airflow obstruction revisited: An analysis of the Framingham Offspring Cohort. American Journal of Respiratory and Critical Care Medicine 2009;180(1):3-10. doi: 10.1164/rccm. 200901-0047OC.

[13] Laniado-Laborin R. Smoking and chronic obstructive pulmonary disease (COPD). Parallel epidemics of the 21st century. International Journal of Environmental Research and Public Health 2009;6(1):209–224. doi: 10.3390/ ijerph6010209.

[14] Rodriguez J, Jiang R, Johnson WC, MacKenzie BA, Smith LJ, Barr RG. The association of pipe and cigar use with cotinine levels, lung function, and airflow obstruction: A cross-sectional study. Annals of Internal Medicine 2010;152(4):201-210. doi: 10.7326/0003-4819-152-4-201002160-00004.

[15] Raad D, Gaddam S, Schunemann HJ, Irani J, Abou Jaoude P, Honeine R, et al. Effects of water-pipe smoking on lung function: A systematic review and meta-analysis. Chest 2011;139(4):764-774. doi: 10.1378/chest.10-0991.

[16] Tan WC, Bsc C Lo, Jong A, Xing L, Gerald MJF, Vollmer WM, et al. Marijuana and chronic obstructive lung disease: A population-based study. CMAJ 2009;180(8):814–820. doi: 10.1503/cmaj.081040.

[17] Yin P, Jiang C, Cheng K, Lam T, Lam K, Miller M, et al. Passive smoking exposure and risk of COPD among adults in China: the Guangzhou Biobank Cohort Study. Lancet 2007;370(9589):751-757. doi: 10.1016/S0140-6736(07) 61378-6.

[18] Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D, et al. An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 2010; 182(5):693-718. doi: 10.1164/rccm.200811-1757ST.

[19] Paulin LM, Diette GB, Blanc PD, Putcha N, Eisner MD, Kanner RE, et al. Occupational exposures are associated with worse morbidity in patients with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 2015;191(5):557-565. doi: 10.1164/rccm.201408-1407 OC.

[20] Marchetti N, Garshick E, Kinney GL, McKenzie A, Stinson D, Lutz SM, et al. Association between occupational exposure and lung function, respiratory symptoms, and high-resolution computed tomography imaging in COPDGene. American Journal of Respiratory and Critical Care Medicine 2014;190(7):756-762. doi: 10.1164/rccm.201403-0493OC.

[21] Raherison C, Girodet PO. Epidemiology of COPD. European Respiratory Review 2009;18:213-221. doi: 10.1183/09059180.00003609.

[22] Mercado N, Ito K, Barnes PJ. Accelerated ageing of the lung in COPD: New concepts. Thorax 2015;70(5):482-489. doi: 10.1136/thoraxjnl-2014-206084.

[23] Landis SH, Muellerova H, Mannino DM, Menezes AM, Han MLK, van der Molen T, et al. Continuing to confront COPD international patient survey: Methods, COPD prevalence, and disease burden in 2012-2013. International Journal of Chronic Obstructive Pulmonary Disease 2014;9:597-611. doi: 10.2147/COPD.S61854.

[24] Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, et al. Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene study. American Journal of Respiratory and Critical Care Medicine 2011; 184(4):414-420. doi: 10.1164/rccm.201011-1928OC.

[25] Lopez Varela MV, Montes De Oca M, Halbert RJ, Muiño A, Perez-Padilla R, Tálamo C, et al. Sex-related differences in COPD in five Latin American cities: The PLATINO study. The European Respiratory Journal 2010; 36(5):1034-1041. doi: 10.1183/09031936.00165409.

[26] Silverman EK, Weiss ST, Drazen JM, Chapman HA, Carey V, Campbell EJ, et al. Gender-related differences in severe, early-onset chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 2000;162(6):2152-2158. doi: 10.1164/ajrccm.162.6.2003112.

[27] Silva GE, Sherrill DL, Guerra S, Barbee RA. Asthma as a risk factor for COPD in a longitudinal study. Chest 2004;126(1):59-65. doi: 10.1378/chest.126.1.59.

[28] Vonk JM, Jongepier H, Panhuysen CIM, Schouten JP, Bleecker ER, Postma DS. Risk factors associated with the presence of irreversible airflow limitation and reduced transfer coefficient in patients with asthma after 26 years of follow up. Thorax 2003;58(4):322-327. doi: 10.1136/thorax.58.4.322.

[29] Barker DJP, Godfrey KM, Fall C, Osmond C, Winter PD, Shaheen SO. Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease. British Medical Journal 1991;303(6804):671–675. doi: 10.1136/bmj.303.6804.671.

[30] Todisco T, de Benedictis FM, Iannacci L, Baglioni S, Eslami A, Todisco E, et al. Mild prematurity and respiratory functions. European Journal of Pediatrics 1993;152(1):55-58. doi: 10.1007/bf02072517.

[31] Allinson JP, Hardy R, Donaldson GC, Shaheen SO, Kuh D, Wedzicha JA. Combined impact of smoking and early-life exposures on adult lung function trajectories. American Journal of Respiratory and Critical Care Medicine 2017; 196(8):1021-1030. doi: 10.1164/rccm.201703-0506OC.

[32] Stoller JK, Aboussouan LS. A review of α 1-antitrypsin deficiency. American Journal of Respiratory and Critical Care Medicine 2012;185(3):246-259. doi: 10.1164/rccm.201108-1428CI.

[33] McCloskey SC, Patel BD, Hinchliffe SJ, Reid ED, Wareham NJ, Lomas DA. Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. American Journal of Respiratory and Critical Care Medicine 2001;164(8 Pt 1):1419-1424. doi: 10.1164/ajrccm.164. 8.2105002.

[34] Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. The New England Journal of Medicine 2004;350(26):2645-2653.

[35] Williams N. The MRC breathlessness scale. Occupational medicine 2017 67(6):496-497. doi: 10.1093/occmed/kqx086.

[36] Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. The European Respiratory Journal 2009;34(3):648-654. doi: 10.1183/09031936.00102509.

[37] Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. The New England Journal of Medicine 2010;363(12):1128-1138. doi: 10.1056/NEJMoa0909883.

[38] Wedzicha JA, Seemungal TA. COPD exacerbations: defining their cause and prevention. Lancet 2007;370(9589):786-796. doi: 10.1016/S0140-6736(07) 61382-8.

[39] Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. Chest 2000;117(5 Suppl 2):398S-401S. doi: 10.1378/chest.117.5_suppl_2. 398s.

[40] Seemungal TAR, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 1998;157(5 Pt 1):1418-1422. doi: 10.1164/ajrccm.157.5.9709032.

[41] Seemungal TAR, Hurst JR, Wedzicha JA. Exacerbation rate, health status and mortality in COPD – a review of potential interventions. International Journal of Chronic Obstructive Pulmonary Disease 2009;4:203–223. doi: 10.2147/ copd.s3385.

[42] Woodhead M, Blasi F, Ewig S, Garau J, Huchon G, Ieven M, et al. Guidelines for the management of adult lower respiratory tract infections - Summary. Clinical Microbiology and Infection 2011; 17 Suppl 6:1-24. doi: 10.1111/j. 1469-0691.2011.03602.x.

[43] Hegele RG, Hayashi S, Hogg JC, Paré PD. Mechanisms of airway narrowing and hyperresponsiveness in viral respiratory tract infections. American Journal of Respiratory and Critical Care Medicine 1995;151(5):1659-1664. doi: 10.1164/ajrccm.151.5.7735630.

[44] Smith CB, Kanner RE, Golden CA, Klauber MR, Renzetti AD. Effect of viral infections on pulmonary function in patients with chronic obstructive pulmonary diseases. Journal of Infectious Diseases 1980;141(3):271-80. doi: 10.1093/infdis/141.3.271.

[45] Seemungal TAR, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time Course and Recovery of Exacerbations in Patients with Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine 2000;161(5):1608-1613. doi: 10.1164/ajrccm.161.5.9908022.

[46] Chang C, Guo Z, Shen N, He B, Yao W, Zhu H, et al. Dynamics of inflammation resolution and symptom recovery during AECOPD treatment. Scientific Reports 2014;4:5516. doi: 10.1038/srep05516.

[47] Fletcher C, Peto R. The natural history of chronic airflow obstruction. The British Medical Journal 1977;1(6077):1645–1648. doi: 10.1136/bmj.1.6077. 1645.

[48] Wongsurakiat P, Maranetra KN, Wasi C, Kositanont U, Dejsomritrutai W, Charoenratanakul S. Acute respiratory illness in patients with COPD and the effectiveness of influenza vaccination: A randomized controlled study. Chest 2004;125(6):2011-2020. doi: 10.1378/chest.125.6.2011.

[49] Poole P, Chacko EE, Wood-Baker R, Cates CJ. Influenza vaccine for patients with chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2006;1:CD002733. doi: 10.1002/14651858.CD002733.pub2.

[50] Nichol KL, Margolis KL, Wuorenma J, Von Sternberg T. The efficacy and cost effectiveness of vaccination against influenza among elderly persons living in the community. The New England Journal of Medicine 1994;331(12):778-784. doi: 10.1056/NEJM199409223311206.

[51] Walters JA, Tang JNQ, Poole P, Wood-Baker R. Pneumococcal vaccines for preventing pneumonia in chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2017;1:CD001390. doi: 10.1002/14651858. CD001390.pub4.

[52] Ejiofor S, Turner AM. Pharmacotherapies for COPD. Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine 2013;7:17–34. doi: 10.4137/CCRPM.S7211.

[53] Cranston JM CAMJ, Alpers JH. Domiciliary oxygen for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2005;4:CD001744. doi: 10.1002/14651858.CD001744.pub2.

[54] Agusti À, Soriano JB. COPD as a systemic disease. COPD: Journal of Chronic Obstructive Pulmonary Disease 2008;5(2):133-138. doi: 10.1080/15412550801941349.

[55] Sinden NJ, Stockley RA. Systemic inflammation and comorbidity in COPD: A result of "overspill" of inflammatory mediators from the lungs? Review of the evidence. Thorax 2010;65(10):930-936. doi: 10.1136/thx.2009.130260.

[56] Morgan AD, Zakeri R, Quint JK. Defining the relationship between COPD and CVD: what are the implications for clinical practice? Therapeutic Advances in Respiratory Disease 2018;12:1753465817750524. doi: 10.1177/1753465817750524.

[57] Durham AL, Adcock IM. The relationship between COPD and lung cancer. Lung Cancer 2015;90(2): 121–127. doi: 10.1016/j.lungcan.2015.08.017.

[58] Ho TW, Huang CT, Ruan SY, Tsai YJ, Lai F, Yu CJ. Diabetes mellitus in patients with chronic obstructive pulmonary disease-The impact on mortality. PLoS One 2017;12(4):e0175794. doi: 10.1371/journal.pone.0175794.

[59] Inoue D, Watanabe R, Okazaki R. COPD and osteoporosis: Links, risks, and treatment challenges. International Journal of Chronic Obstructive Pulmonary Disease 2016;11:637-648. doi: 10.2147/COPD.S79638.

[60] Vestbo J, Prescott E, Almdal T, Dahl M, Nordestgaard BG, Andersen T, et al. Body mass, fat-free body mass, and prognosis in patients with chronic obstructive pulmonary disease from a random population sample: Findings from the Copenhagen City Heart Study. American Journal of Respiratory and Critical Care Medicine 2006;173(1):79-83. doi: 10.1164/rccm.200506-969OC.

[61] Incalzi RA, Corsonello A, Pedone C, Battaglia S, Paglino G, Bellia V. Chronic Renal Failure: A Neglected Comorbidity of COPD. Chest 2010;137(4):831-837. doi: 10.1378/chest.09-1710.

[62] Yohannes AM, Alexopoulos GS. Depression and anxiety in patients with COPD. European Respiratory Review 2014;23:345-349. doi: 10.1183/09059180.00007813.

[63] Franssen FME, Rochester CL. Comorbidities in patients with COPD and pulmonary rehabilitation: Do they matter? European Respiratory Review 2014;23(131):131-141. doi: 10.1183/09059180.00007613.

[64] Baty F, Putora PM, Isenring B, Blum T, Brutsche M. Comorbidities and Burden of COPD: A Population Based Case-Control Study. PLoS One 2013;8(5):e63285. doi: 10.1371/journal.pone.0063285.

[65] Burgel PRG, Roche N, Paillasseur JL, Tillie-Leblond I, Chanez P, Escamillae R, et al. Clinical COPD phenotypes identified by cluster analysis: Validation with mortality. The European Respiratory Journal 2012;40(2):495-496. doi: 10.1183/09031936.00228511.

[66] Malo de Molina R, Aguado S, Arellano C, Valle M, Ussetti P. Ischemic Heart Disease during Acute Exacerbations of COPD. Medical Sciences 2018;6(4): 83. doi: 10.3390/medsci6040083.

[67] Han MLK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, et al. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: Associated radiologic phenotypes. Radiology 2011;261(1):274-282. doi: 10.1148/radiol.11110173.

[68] Kim V, Han MLK, Vance GB, Make BJ, Newell JD, Hokanson JE, et al. The chronic bronchitic phenotype of COPD: An analysis of the COPDGene study. Chest 2011;140(3):626-633. doi: 10.1378/chest.10-2948.

[69] Connors AF, Dawson N V., Thomas C, Harrell FE, Desbiens N, Fulkerson WJ, et al. Outcomes following acute exacerbation of severe chronic obstructive lung disease. American Journal of Respiratory and Critical Care Medicine 1996;154(4 Pt 1):959-967. doi: 10.1164/ajrccm.154.4.8887592.

[70] Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 2004;169(12):1298-1303. doi: 10.1164/rccm.200310-1443OC.

[71] Langsetmo L, Platt RW, Ernst P, Bourbeau J. Underreporting exacerbation of chronic obstructive pulmonary disease in a longitudinal cohort. American Journal of Respiratory and Critical Care Medicine 2008;177(4):396-401. doi: 10.1164/rccm.200708-1290OC.

[72] Xu W, Collet JP, Shapiro S, Lin Y, Yang T, Wang C, et al. Negative impacts of unreported COPD exacerbations on health-related quality of life at 1 year. The European Respiratory Journal 2010;35(5):1022-1030. doi: 10.1183/09031936.00079409.

[73] Jones PW, Lamarca R, Chuecos F, Singh D, Agustí A, Bateman ED, et al. Characterisation and impact of reported and unreported exacerbations: Results from ATTAIN. The European Respiratory Journal 2014;44(5):1156-1165. doi: 10.1183/09031936.00038814.

[74] Anzueto A. Impact of exacerbations on COPD. European Respiratory Review 2010;19(116):113-118. doi: 10.1183/09059180.00002610.

[75] McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2011;7:CD007718. doi: 10.1002/14651858.CD007718.pub2.

[76] Trappenburg JCA, Niesink A, De Weert-Van Oene GH, Van Der Zeijden H, Van Snippenburg R, Peters A, et al. Effects of telemonitoring in patients with chronic obstructive pulmonary disease. Telemedicine Journal and e-Health. 2008;14(2):138-146. doi: 10.1089/tmj.2007.0037.

[77] Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: A systematic review and meta-analysis. Journal of Telemedicine and Telecare. 2010;16(3):120-127. doi: 10.1258/jtt.2009.090812.

[78] Berkhof FF, Van Den Berg JWK, Uil SM, Kerstjens HAM. Telemedicine, the effect of nurse-initiated telephone follow up, on health status and health-care utilization in COPD patients: A randomized trial. Respirology 2015;20(2):279-285. doi: 10.1111/resp.12437.

[79] McDowell JE, McClean S, FitzGibbon F, Tate S. A randomised clinical trial of the effectiveness of home-based health care with telemonitoring in patients with COPD. Journal of Telemedicine and Telecare 2015;21(2):80-87. doi: 10.1177/1357633X14566575.

[80] Ringbæk T, Green A, Laursen LC, Frausing E, Brøndum E, Ulrik CS. Effect of tele health care on exacerbations and hospital admissions in patients with chronic obstructive pulmonary disease: A randomized clinical trial. International Journal of Chronic Obstructive Pulmonary Disease 2015;10: 1801–1808. doi: 10.2147/COPD.S85596.

[81] Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in chronic obstructive pulmonary disease (chromed) a randomized clinical trial. American Journal of Respiratory and Critical Care Medicine 2018;198(5):620-628. doi: 10.1164/rccm.201712-2404OC.

[82] Kruse C, Pesek B, Anderson M, Brennan K, Comfort H. Telemonitoring to Manage Chronic Obstructive Pulmonary Disease: Systematic Literature Review. JMIR Medical Informatics 2019;7(1):e11496. doi: 10.2196/11496.

[83] Sanchez-Morillo D, Fernandez-Granero MA, Leon-Jimenez A. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: A systematic review. Chronic Respiratory Disease 2016;13(3):264-283. doi: 10.1177/1479972316642365.

[84] Bishop CM. Pattern Recognition and Machine Learning. New York: Springer 2006.

[85] Polikar R. Pattern Recognition. Wiley Encyclopedia of Biomedical Engineering 2006. doi: 10.1002/9780471740360.ebs0904.

[86] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006;27(8):861-874.

[87] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29–36. doi: 10.1148/radiology.143.1.7063747.

[88] Jensen MH, Cichosz SL, Dinesen B, Hejlesen OK. Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. Journal of Telemedicine and Telecare 2012;18(2):99–103. doi: 10.1258/jtt. 2011.110607.

[89] van der Heijden M, Lucas PJF, Lijnse B, Heijdra YF, Schermer TRJ. An autonomous mobile system for the management of COPD. Journal of Biomedical Informatics 2013;46(3):458-469. doi: 10.1016/j.jbi.2013.03.003.

[90] Fernandez-Granero MA, Sanchez-Morillo D, Leon-Jimenez A, Crespo LF. Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms. Bio-medical Materials and Engineering 2014;24(6):3825–3832. doi: 10.3233/BME-141212.

[91] Mohktar MS, Redmond SJ, Antoniades NC, Rochford PD, Pretto JJ, Basilakis J, et al. Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. Artificial Intelligence in Medicine 2015;63(1):51–59. doi: 10.1016/j.artmed.2014.12.003.

[92] Sanchez-Morillo D, Fernandez-Granero MA, Jiménez AL. Detecting COPD exacerbations early using daily telemonitoring of symptoms and k-means clustering: a pilot study. Medical & Biological Engineering & Computing 2015;53(5):441-451. doi: 10.1007/s11517-015-1252-4.

[93] Riis HC, Jensen MH, Cichosz SL, Hejlesen OK. Prediction of exacerbation onset in chronic obstructive pulmonary disease patients. Journal of Medical Engineering & Technology 2016;40(1):1-7. doi: 10.3109/03091902.2015.1105317.

[94] Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: Identification and prediction using a digital health system. Journal of Medical Internet Research 2017;19(3):e69. doi: 10.2196/jmir.7207.

[95] Sadatsafavi M, Xie H, Etminan M, Johnson K, FitzGerald JM. The association between previous and future severe exacerbations of chronic obstructive pulmonary disease: Updating the literature using robust statistical methodology. PLoS One 2018;13(1):e0191243. doi: 10.1371/journal.pone.0191243.

[96] Urwyler P, Abu Hussein N, Bridevaux PO, Chhajed PN, Geiser T, Grendelmeier P, et al. Predictive factors for exacerbation and re-exacerbation in chronic obstructive pulmonary disease: An extension of the Cox model to analyze data from the Swiss COPD cohort. Multidisciplinary Respiratory Medicine 2019;14:7. doi: 10.1186/s40248-019-0168-5.

[97] Jenkins CR, Celli B, Anderson JA, Ferguson GT, Jones PW, Vestbo J, et al. Seasonality and determinants of moderate and severe COPD exacerbations in the TORCH study. The European Respiratory Journal 2012;39(1):38-45. doi: 10.1183/09031936.00194610.

[98] Lopez-Campos JL, Calero C, Quintana-Gallego E. Symptom variability in COPD: A narrative review. International Journal of Chronic Obstructive Pulmonary Disease 2013;8:231-238. doi: 10.2147/COPD.S42866.

[99] Donaldson GC, Wedzicha JA. The causes and consequences of seasonal variation in COPD exacerbations. International Journal of Chronic Obstructive Pulmonary Disease 2014;9:1101–1110. doi: 10.2147/COPD.S54475.

[100] Li J, Sun S, Tang R, Qiu H, Huang Q, Mason TG, et al. Major air pollutants and risk of COPD exacerbations: A systematic review and meta-analysis. International Journal of Chronic Obstructive Pulmonary Disease 2016;11: 3079–3091. doi: 10.2147/COPD.S122282.

[101] ClinicalTrials.gov. Telemedicine for Patients Suffering From COPD (Danish Telecare North Trial) (TCN) 2013. Identifier: NCT01984840.

[102] Mylund L. Telemedicin hos patienter med kronisk obstruktiv lungesygdom (Danish) 2012. Accessed November 2019. Available from: https://www.yumpu.com/da/document/read/21663667/telemedicin-hos-patienter-med-kronisk-obstruktiv-lungesygdom/7.

[103] Halpin DMG, Miravitlles M, Metzdorf N, Celli B. Impact and prevention of severe exacerbations of COPD: A review of the evidence. International Journal of Chronic Obstructive Pulmonary Disease 2017;12:2891-2908. doi: 10.2147/COPD.S139470.

[104] Pavord ID, Jones PW, Burgel P-R, Rabe KF. Exacerbations of COPD. International Journal of Chronic Obstructive Pulmonary Disease 2016;11(Spec Iss):21–30. doi: 10.2147/COPD.S85978.

[105] Danish Agency for Digitisation, Ministry of Finance Denmark. New national strategy: Artificial intelligence should benefit individuals, businesses and society as a whole. Accessed December 2019. Available from: https://en.digst.dk/news/news-archive/2019/march/new-national-strategy-artificial-intelligence-should-benefit-individuals-businesses-and-society-as-a-whole/.

ISSN (online): 2246-1302
ISBN (online): 978-87-7210-566-6

AALBORG UNIVERSITY PRESS