Teaching Pepper Robot to Recognize Emotions of Traumatic Brain Injured Patients Using Deep Neural Networks

Ilyas, Chaudhary Muhammad Aqdus; Schmuck, Viktor; Haque, Mohammad Ahsanul; Nasrollahi, Kamal; Rehm, Matthias; Moeslund, Thomas B.

Published in:
28th IEEE International Conference on Robot and Human Interactive Communication (ROMAN)

DOI (link to publication from Publisher):
10.1109/RO-MAN46459.2019.8956445

Publication date:
2019

Document Version
Early version, also known as pre-print

Teaching Pepper Robot to Recognize Emotions of Traumatic Brain Injured Patients Using Deep Neural Networks

Chaudhary Muhammad Aqdus Ilyas[1] Viktor Schmuck[2] Muhammad Ahsanul Haque[3]
Kamal Nasrollahi[4] Matthias Rehm[5] and Thomas B. Moeslund[6]

***Abstract*— Social signal extraction from the facial analysis is a popular research area in human-robot interaction. However, recognition of emotional signals from Traumatic Brain Injured (TBI) patients with the help of robots and non-intrusive sensors is yet to be explored. Existing robots have limited abilities to automatically identify human emotions and respond accordingly. Their interaction with TBI patients could be even more challenging and complex due to unique, unusual and diverse ways of expressing their emotions. To tackle the disparity in a TBI patient's Facial Expressions (FEs), a specialized deep-trained model for automatic detection of TBI patients' emotions and FE (TBI-FER model) is designed, for robot-assisted rehabilitation activities. In addition, the Pepper robot's built-in model for FE is investigated on TBI patients as well as on healthy people. Variance in their emotional expressions is determined by comparative studies. It is observed that the customized trained system is highly essential for the deployment of Pepper robot as a Socially Assistive Robot (SAR).**

I. INTRODUCTION

Researchers have conducted extensive investigation into human-focused robotic technologies, designed to achieve real time and close to human-like human-robot interactions [1]. However, existing robotic technologies that facilitate robots in human emotions recognition have limitations [1] and require more intelligent platforms and software to communicate and respond naturally with people [2]. Recently robots have been developed to collaborate with doctors, physicians or physiotherapists. In the health care sector these robots are tailored-made, particularly Socially Assistive Robots (SAR), to provide assistance and improvement in a wide range of medical applications such as robot-assisted therapies [3], [4], complex-surgical operations [5], [6], or for social engagement with people with special needs like children with autism spectrum disorder (ASD) [7]–[10]. Machine learning, especially deep learning, approaches have enabled these robots to automatically identify and react intelligently to subject emotional states. These smart machines require techniques that can accurately and robustly recognize human emotional clues from uncontrolled and natural environmental conditions [11].

A typical robot for health monitoring and improvement needs to receive audio, video or proximity information from its sensors. This information is then processed based on the algorithm that interpret the information into meaningful signals. This is followed with robot action or response for the desired task [7]. In some cases, therapist or 'an agent behind the curtain' controls the robots due to lack of automatic perception of signals and spontaneous response to the emotional cues, consequently making less autonomous human-robot interaction. There is a need of autonomous and data-driven machines that can determine patient behavior and react accordingly [12]. Furthermore, these systems are heavily relying on both audio and video sensors input for making stronger relation. However, robots placement to aid TBI patients in a home or in a specialized neuro-center, face certain additional obstacles that are necessary to be considered. These include the patients' non-cooperative behavior, inappropriate responses and inability to express their emotions. This is due to the nature of the condition like stroke or accident, resulting in damaged sensory motor control and reasoning skills, along with restricted muscle movements due to paralysis [13], [14]. However, these challenges can be different from patient-to-patient, depending on the nature and severity of the injury, producing speech inhibition, partial or complete paralysis, involuntary body movements, abrupt emotional changes, aggression, lack of consciousness or attention and varied emotion elicitation [15]. Therefore, we aim to exploit only visual signals for system generalization for TBI patient's emotional analysis through facial expressions.

Current Facial Expression Recognition (FER) systems are largely based on Convolutional Neural Networks (CNN) for feature extraction and classification as they provide state-of-art results for face recognition [16], [17], facial expression recognition [18]–[21] and emotional states identification [22], [23]. Their results are highly accurate on healthy people and in controlled conditions. However, this high accuracy is still yet to be achieved with challenging environmental conditions such as large pose variation, low illumination, and on data sets of people with limited expressions like TBI patients. In addition to that, remarkable achievements has witnessed in machines analysis of human emotions, but there are still noticeable challenges that are needed to be addressed in order to involve robots into daily interfaces like social, physical or cognitive activities in real-world scenarios. Some of the major challenges are as follows:

- The wide range of datasets available for FER are collected in laboratory and controlled conditions with little or no pose variations, frontal views, without occlusion, stable illumination and with cooperative subjects. Undoubtedly, such luxuries are not present in real-

[1] [3] [4] [6] Dept. of Architecture, Design and Media Technology, (Visual Analysis of People (VAP), Aalborg University, Denmark cmai,mah,kn,tbm@create.aau.dk
[2] [5] Dept. of Architecture, Design and Media Technology, (Interaction Laboratory (IL), Aalborg University, Denmark vsch,matthias@create.aau.dk

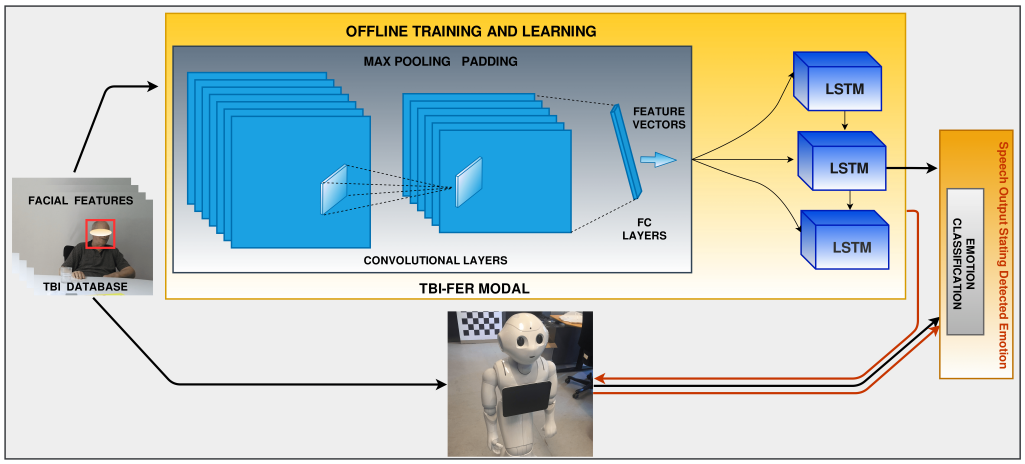

Fig. 1. Block Diagram of Deep Trained TBI-FER Model and Pepper's Built-in Facial Expression Recognition for Emotion Classification. Black-arrow represents FER through built-in Pepper robot whereas red-arrows represents deployment of TBI-FER modal

world applications. Systems trained on such data do not perform well in real-time with real subjects.

- Currently available datasets have FE of healthy people who are mostly cooperative and sometimes produce induced-expressions as compared to TBI patients who have impaired skills, and quite varied and limited expressions due to facial paralysis [13], [14]. Additionally, induced FEs that are produced consciously largely alter from natural involuntary emotions.
- Most of existing intelligent systems are trained on databases where expressions are clear with little variance on the vast majority of all 6 basic expressions such as happiness, sadness, fear, angry, surprise and disgust. However, in case of TBI patients classification of 6 basic expressions (7 including neutral) is quite complex as TBI patient's expressions are not easily distinguishable except for one or two and all expressions are not very common. Hence, SARs trained on these databases needed to be customized as these special subjects behave and respond differently than healthy people.

In this research article, we intend to address the aforementioned complexities and limitations in TBI-human-robot interaction by the utilization of a TBI-patient database, which is a collection of multimodal data annotated by TBI-patient's care givers, experts, physiotherapists and doctors. This database is a collection of TBI facial images for spontaneous expression analysis, captured in an entirely unconstrained, real-world environment. It contains the events of natural interactions of subjects of diverse background and age groups in three scenarios of cognitive, social and physical rehabilitation activities. We used this database to develop a deep trained model (TBI-FER Model), composed of Convolutional Neural Network and Long Short Term Memory Networks (CNN-LSTM) to exploit the spatio-temporal information of the TBI subjects. This TBI-FER Model is dedicated for FER of TBI patients that can be integrated with SAR robot, like the Pepper robot for effective human-robot-engagement-research. We performed the classification of 6 basic expressions through the TBI-FER model validated on the TBI patient database as well as the Extended Cohn-Kanade (CK+) (healthy people) database [24]. We also present the hypothesis that our proposed model will outperform the pepper robot built in model. Furthermore, the Pepper robot built-in FER model is employed on both healthy and TBI patient databases and a FE variance analysis is made.

The rest of this article is structured as follows. Section II presents the related work including social robots and facial expression recognition. Section III describes the methodology for the TBI-FER model training with CNN-LSTM and FE identification through the Pepper robot, and Section IV describes the experiment and its results. Finally, Section V presents the discussion and concludes the paper.

II. Related Work

A. Social Assistive Robots

In recent years, there has been a growing interest in providing assistance and services to people for physical or cognitive rehabilitation, social interactions and many other health care applications with the help of special robots, categorized as social assistive robots (SAR) [25]. SAR are extensively purposed for monitoring and assisting elderly people in activities of daily life (ADL) in smart homes. Paro, a pet robot resembling a baby seal, has shown positive results for pet therapy by reducing stress in residents in care centers [26]. This also resulted in increased social interaction between residents. Similarly, Roball, a mobile SAR with an IR sensor for touch detection, improved the social interactive behaviour among kids suffering from Autism Spectrum Disorders (ASDs). Roball has encouraged the kids to play with trainers, therapists, and family members [27]. AIBO (Artificial Intelligence roBOt), an autonomous entertainment robot, was proved effective in enhancing social interaction as well as in aiding mental therapy [28]. AIBO uses touch, audio, vision and thermal sensors to perceive information. A

personal assistant robot, Philips iCat, used as a companion, motivator and educator, performed roles of engaging, fostering and instructing [29]. iCat uses vocal emotional expression as well as facial emotional expressions. Another type of robot architecture that integrates the domains of robotics, medicine, psychology, social, cognitive as well as interactive fields is HealthBot [30]. This robot was designed to help the elderly, monitoring their health status and detecting falls. In addition, there is an extensive research on assistive and interactive robotics focusing on the rehabilitation of the elderly and people who suffered stroke [31], [32]. The mentioned companion robots aid in ADL [33] and engaging socially for the purpose of assistance and recovery to improve life quality [30], [34], [35] in the field as well as in lab. Sophia, one of the most advanced humanoid social robot can display expressions similar to humans to build trust and aid humans towards a better life and design smarter homes [36]. She has the ability to process visual, emotional and conversational data. Sophia incorporates Gardner's multiple intelligences [37] into her cognitive architecture. Sophia has also been used as a meditation consultant, giving step by step instructions to help people feel better in lab environment but Sophia has not been placed in field with real subjects. Additionally, these robots utilize different perceived signals such as voice, touch, gestures, signals through IR, RGB, thermal and depth cameras, subject motion tracking, force sensors, and many other indicators to perform their tasks.

Mabu, the intelligent and socially interactive personal health care companion, looks after the patients at home, and mainly reminds them about their medication [38]. Mabu emotionally engages with patients, and evolves its relationship over time by tailoring its conversation by adopting behavior psychology using Artificial Intelligence (AI) algorithms [39]. It also focuses on keeping the patients healthy by constantly monitoring their health and sending encrypted data to a personal doctor if required. Moreover, it actively involves its patients in therapies as prescribed by the doctors. One of the major features of Mabu is active involvement in its speech with patients and the ability to augment its psychological and physiological models to generate new conversational models with the aim of long-term health care [38], [39].

SoftBank robotics have developed NAO [40] and Pepper [41], which are high performance humanoid robots for research and education purposes with the ability to process a wide range of expressions and gesture information. Pepper is equipped with several sensors, but most importantly two 2D and one 3D cameras, which can easily be accessed by its SDKs. Due to its cameras and sensors the Pepper robot can recognize, track and turn while following faces. It also has a preset FER algorithm. The comparison of the discussed robot's input modalities and re-learning capabilities is presented in Table I whereas their illustration is presented in Figure 2.

TABLE I
ROBOTS' INPUT MODALITIES AND RE-LEARNING CAPABILITY

Robot	Audio Input	Video Input	Tactile Input	Adaptive Re-learning
Sophia [36]	Yes	Yes	No	Yes
Mabu [38]	Yes	Yes	Via tablet	Yes
Pepper [41]	Yes	Yes	Yes	Not by default
NAO [40]	Yes	Yes	Yes	No
iCAT [29]	Yes	Yes	No	No
HealthBot [30]	Yes	No	Via tablet	No
PARO [26]	Yes	Via light sensor	Yes	No
AIBO [28]	Yes	Yes	Yes	No

B. Deep Learning Approaches for Facial Expression Recognition

In the aforementioned robots, different sensors have been integrated to achieve efficient human-robot interaction but in our case we would like to rely only on visual information so that the robot can communicate and recognize the emotions of TBI patients effectively, regardless of their speech and locomotion disabilities. It is observed that human emotions are mostly recognized by facial expressions [42], [43]. In order to identify emotions accurately, face and Facial Expression Recognition (FER) approaches have been evolved from holistic, local-feature-based like Gabor or Local Binary Pattern (LBP), learning-based-local descriptors (shallow methods) to deep learning (DL) methods [44]. Traditional methods failed to address certain challenges when researchers moved towards automated and unconstrained FER in challenging conditions. In 1990's, the holistic approaches dominated the FR community with certain low-dimensional representations inferences like linear subspace, sparse representation and manifold approaches [45], [46]. However, these holistic methods failed when exposed to uncontrolled facial changes, different from prior assumptions. This lead to rise of local features based facial recognition methods involving Local Binary pattern (LBP) [47], Gabor [48], SIFT, HOG and other high-level dimensional representations [49]. Unfortunately, these handcrafted features could not address the unique characteristics and denseness of facial features. Following these limitations, researchers introduced the learning-based-local descriptors for better distinctiveness and compactness. This produced FE accuracy of approximately 95% [43] but this is achieved under controlled conditions with frontal views and high resolution images. However, these shallow methods do not handle well non-linear changes in facial appearance. In real time scenarios, shallow methods have improved the accuracy on challenging unconstrained Labeled Faces in the Wild (LFW) dataset [50] to about 95% [51] in 2010. Alex Net won the Image-Net competition [52], through deep learning methods, such as convolution neural networks (CNNs) with a substantial margins. Similarly, in 2014, DeepFace approached close to human performance (97.53%) on LFW dataset benchmark [50], and acquired

Sophia Mabu Pepper NAO iCat HealthBot (top) PARO (bottom) AIBO

Fig. 2. Famous SAR robots

Fig. 3. Pepper robotic administration for Facial Expression Analysis

state-of-arts performance (97.35%) [18]. All of these experimental evaluations are based on subjects without any expression impairments like TBI patients. Ilyas et al in [13] have exercised the CNN-LSTM architecture to exploit the spatio-temporal information for features classification and mood analysis of TBI patients and achieved an accuracy of 87.97% on challenging TBI database. We have employed the same linear combination of CNN-LSTM architecture to train the TBI database and compared with Pepper robot built-in FER model to have FER performance analysis.

III. Methodology

A. Database Development and Training

The main aim of this study is to perform facial expression (FE) and mood recognition of TBI patients in order to enhance the social interaction and assist trainers and physiotherapists with the help of robots. First we accumulated a database in three uniform scenarios namely cognitive, physio and social rehabilitation activities, ensuring the reliability of the database as explained in detail in [13], [14]. This database is comprised of 924 videos taken about 11 participants, each being a maximum of 5 second in length, recorded with an Axis RGB and a Logitech RGB camera during multiple sessions at 30fps, resulting in approximately 140,000 captured frames.

For database training, first various pre-processing techniques like face detection, landmarks detection and tracking by Supervised Decent Method (SDM) followed by Face Quality Assessment (FQA), were applied to guarantee high quality images in Face-Log system. In the next step, this high quality image database is passed through a linear architecture of CNN and LSTM, to extract the facial features with the help of CNN from the input faces of TBI patients and then feed to LSTM to exploit the temporal relation on the basis of extracted features in timely manner. For feature extraction we have fine tuned the CNN with off the shelf pre-trained VGG-16CNN model [53]. Features are obtained as `fc7` layer of CNN with VGG-16 model that is feed into LSTM model to analyze the performance of combined CNN + LSTM deep neural architecture, resulting in TBI patients' FER model (TBI-FER). For performance evaluation the TBI-FER model is validated on the CK+ database. The general schematic of the robotic architecture executed for FER analysis is demonstrated in Figure 1.

In order to analyze the FER through Pepper robot, the solution required two distinct operations. Firstly, the NaoQi Python SDK is used to retrieve video from the 2D camera of the pepper robot with frame rate of 30FPS and the resolution to 320x240px. Secondly, subject ID file is created and participants were then asked to sit in front of the robot and make different facial expressions related to the 6 emotions. Figure 3 is illustrating the procedure of emotion elicitation through Pepper robot.

IV. Experimental Results

In order to present the results, first we explain the experimental setup. In terms of experiments, we evaluate both FER models namely TBI-FER and Pepper-FER models on TBI and CK+ databases for emotion recognition as seen in Figure 4.

A. Experimental Setup

The robot was set up to perform FER with its built-in detection algorithm in order to later annotate the recorded one minute videos and to serve as a base for comparing the built-in method (Pepper-FER) to our proposed model. This model is also validated on both TBI and CK+ databases. Pepper utilizes that trained model for live classification on the robot. In order to compare with the TBI-FER model, a connection is established with robot similar to video recording and images are retrieved. The images were passed onto the loaded

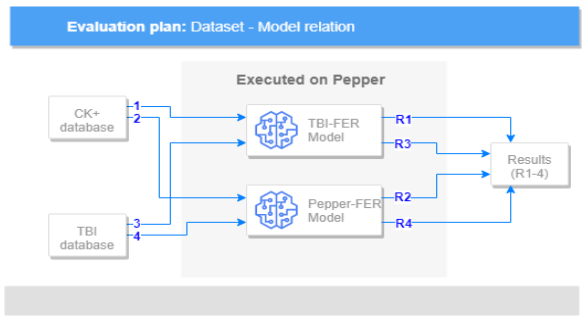

Fig. 4. Evaluation of FER Models on TBI and CK+ Database

TABLE II
CONFUSION MATRIX OF 6 BASIC EXPRESSIONS THROUGH TBI-FER MODEL ON TBI PATIENT'S DATABASE

	Neutral	Happy	Angry	Sad	Fatigued	Surprised
Neutral	**88**	3	2	14	2	1
Happy	4	**82**	2	3	2	7
Angry	2	2	**85**	5	6	1
Sad	12	1	4	**78**	11	1
Fatigued	7	1	5	2	**67**	9
Surprised	2	21	3	2	6	**71**

classification model, and the classified emotion was returned as a string. The information was used to be pushed to the robot through another initialized service converting text to speech (TTS). As a result, the robot was capable of reporting the participants' emotions through TTS with our proposed FER model.

B. TBI-FER Model Analysis

In this section, we discuss the training of our system on the TBI patient database and its validation of the results for 6 basic expressions. It is evident that the neutral expression has the highest, 88% accuracy, as shown in Table II. This is due to the fact that neutral is the most common expression in TBI database. Although, in most cases TBI neutral expression is most likely recognized as sad for healthy people. Fatigued or stress expression exhibits the lowest accuracy in the validation of this FER model. This is due to the unbalanced data set, which is a result of the difficulty of acquiring this type of data because of stressed or non-cooperative participants. On the other hand, when this TBI-FER model is employed on the CK+ database for identification of expressions, it is shown that the CK+ database results are much better compared to the TBI patient one due to the reason that latter database is mainly of high quality images with frontal faces. Comparatively, in case of the TBI patients, there is challenge of working with non-frontal faces. FE of neutral, angry, sad, happy, surprise and fatigue are identified accurately up to 91%, 88%, 87%, 85%, 84% and 82% respectively as illustrated in Table V.

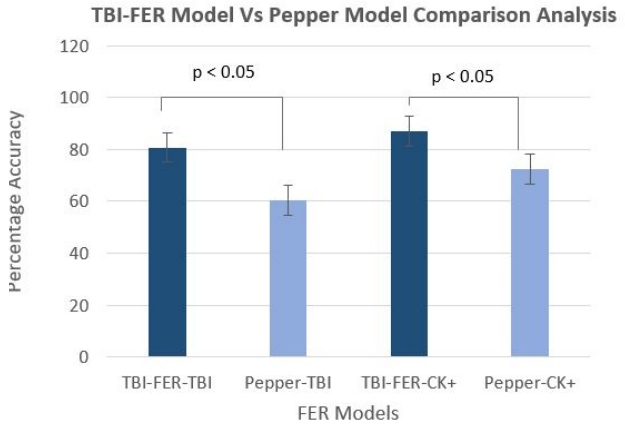

Fig. 5. Comparison of FER Modals on TBI Patients and Healthy Subjects

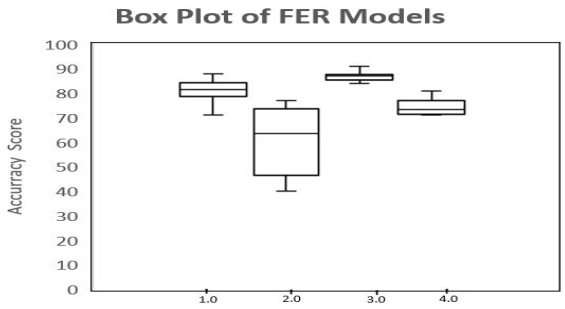

Fig. 6. Box plot of the FER accuracy score. On the x-axis 1.0 is TBI-FER model on the TBI database, 2.0 is Pepper-FER model on the TBI database, 3.0 is TBI-FER model on the CK+ database, 4.0 is Pepper-FER model on the CK+ database

C. Pepper built-in FER Model Analysis

For the classification of emotions through Pepper, its built-in FER model is implemented on both TBI patient and healthy people database. It is observed that Pepper identified the surprise emotion from TBI patients with an accuracy of 42% as opposed to 71% for CK+ database as demonstrated in Table III and IV respectively. This can be due to the varied and limited surprise elicitation from TBI patients due to stroke impact. Furthermore, Pepper identifies neutral expressions of TBI patients with only 42% accuracy with sad and neutral expression overlapping, proving that TBI patients' neutral expressions are more likely recognized as sad ones. Experts have annotated the patients' expressions as neutral since their ability to display emotional signals is disturbed due to injury, and during post stroke rehabilitation they exhibit depression and negative emotions more often than positive ones [13], [14]. It is also observed that the Pepper robot failed to identify fatigue expressions due to technical limitations.

In order to determine which FER model is significantly more accurate, we have conducted a student's t-test on the TBI-FER model and the Pepper built-in model, where variance is approximated for each of the model. For t-test each of the model has to follow the normal distribution and this validated by Q-Q plots and K-S normality tests. We conducted t-tests on two separate databases for each of the FER model. As seen in Figure 5, for TBI database, the t-value comes out 2.54 with a p-value 0.023. Thus, the null hypothesis can be rejected and we can conclude that the TBI-FER model is significantly more accurate. By studying the box plot in Figure 6, it can be seen that TBI-FER score is greater than Pepper-FER score, it can be concluded that TBI-FER model has higher accuracy than the Pepper-FER model on the TBI-database. Similarly, when examining the CK+ database for FER models accuracy, the t-value comes out 3.17 with the p-value 0.003. Thus, we can conclude that the TBI-FER model is also significantly more accurate than the Pepper FER model for healthy subjects. It is also clearly evident in the box plot in Figure 6, the TBI-FER model has higher score than Pepper-FER model on CK+ database.

TABLE III

CONFUSION MATRIX OF FACIAL EXPRESSION RECOGNITION THROUGH PEPPER-ROBOT BUILT-IN MODAL ON TBI PATIENTS

	Neutral	Happy	Angry	Sad	Fatigued	Surprised
Neutral	**42**	0	12	18	x	1
Happy	1	**67**	2	0	x	5
Angry	12	5	**73**	9	x	2
Sad	17	1	12	**76**	x	2
Fatigued*	x	x	x	x	x	0
Surprised	2	2	3	2	x	**42**

* The Pepper robot lacks ability to identify fatigue expressions.

TABLE IV

CONFUSION MATRIX OF FACIAL EXPRESSION RECOGNITION THROUGH PEPPER-ROBOT BUILT-IN MODAL ON HEALTHY PEOPLE

	Neutral	Happy	Angry	Sad	Fatigued	Surprised
Neutral	**59**	1	5	7	x	1
Happy	14	**74**	2	2	x	23
Angry	11	3	**78**	4	x	2
Sad	17	1	9	**81**	x	3
Fatigued*	x	x	x	x	x	0
Surprised	2	15	7	2	x	**71**

* The Pepper robot lacks ability to detect fatigue expressions.

TABLE V

CONFUSION MATRIX OF 6 BASIC EXPRESSIONS THROUGH TBI-FER MODEL ON CK+ DATABASE

	Neutral	Happy	Angry	Sad	Fatigued	Surprised
Neutral	**91**	2	3	5	1	1
Happy	3	**85**	2	3	2	4
Angry	2	2	**88**	5	6	2
Sad	5	1	4	**87**	12	2
Fatigued	5	1	5	3	**82**	2
Surprised	5	4	1	2	6	**84**

V. CONCLUSION AND DISCUSSION

In the general context of FER and social interaction of TBI patients, this paper has presented a robotic framework to identify the FE and emotional signals of TBI patients specifically by introduction of customized deep trained model to meet the requirements of a specialized scenario. To do so, two FER-models, customized TBI-FER model and Pepper-FER model are compared, and their performance is analyzed. For this purpose, TBI patients database was collected in three uniform scenarios, than deep trained model composed of linear combination of CNNs and LSTM is developed to identify the FE and mood of TBI patients. This model is compared with the Pepper robot built-in FER model and FER accuracy is determined using objective assessment methods. Objective evaluation method is used by analyzing facial expressions on test subjects. The results demonstrated that TBI-FER model has significantly higher performance as compared to the Pepper-FER model, on both TBI database and CK+ database (healthy subjects). Furthermore, individual expressions are more pronounced by TBI-FER model, this cross validates the previous results. So in order to place the Pepper robot with TBI patients, it is essential to use customized trained model for more meaningful interaction. Facial expression recognition has proved to be a vital tool to evaluate the mood of subjects in non-obtrusive manner for enhancing social interaction. Therefore, the Pepper robot can use these self-trained models, in our case a TBI-FER model. This can lead to behavioral adaptation of the robot in accordance with patient mood, similar to the implementations of Mabu and Sophia [36], [38] but with less cost and computational power.

REFERENCES

[1] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 2016, pp. 3723–3726.

[2] A. E. Eiben, "Grand challenges for evolutionary robotics," *Frontiers in Robotics and AI*, vol. 1, p. 4, 2014.

[3] D. Feil-Seifer and M. J. Matarić, "Socially assistive robotics," *IEEE Robotics & Automation Magazine*, vol. 18, no. 1, pp. 24–31, 2011.

[4] M. J. Matarić, "Socially assistive robotics: Human augmentation versus automation," *Science Robotics*, vol. 2, no. 4, p. eaam5410, 2017.

[5] B. Davies, "Robotic surgery–a personal view of the past, present and future," *International Journal of Advanced Robotic Systems*, vol. 12, no. 5, p. 54, 2015.

[6] S. P. DiMaio and S. E. Salcudean, "Needle steering and motion planning in soft tissues," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 6, pp. 965–974, 2005.

[7] O. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard, "Personalized machine learning for robot perception of affect and engagement in autism therapy," *Science Robotics*, vol. 3, no. 19, p. eaao6760, 2018.

[8] P. Chevalier, J.-C. Martin, B. Isableu, C. Bazile, and A. Tapus, "Impact of sensory preferences of individuals with autism on the recognition of emotions expressed by two robots, an avatar, and a human," *Autonomous Robots*, vol. 41, no. 3, pp. 613–635, 2017.

[9] P. Chevalier, "Social personalized human-machine interaction for people with autism," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 229–230.

[10] J. Abbasi, "In-home robots improve social skills in children with autism," *Jama*, vol. 320, no. 14, pp. 1425–1425, 2018.

[11] J. Kossaifi, R. Walecki, Y. Panagakis, J. Shen, M. Schmitt, F. Ringeval, J. Han, V. Pandit, B. Schuller, K. Star *et al.*, "Sewa db: A rich database for audio-visual emotion and sentiment research in the wild," *arXiv preprint arXiv:1901.02839*, 2019.

[12] S. Harker, "Applied behavior analysis (aba)," *Encyclopedia of Child Behavior and Development*, pp. 135–138, 2011.

[13] C. M. A. Ilyas, M. A. Haque, M. Rehm, K. Nasrollahi, and T. B. Moeslund, "Facial expression recognition for traumatic brain injured patients," in *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP*, 2018, pp. 522–530.

[14] C. M. A. Ilyas, M. Rehm, K. Nasrollahi, and T. B. Moeslund, "Rehabilitation of traumatic brain injured patients: Patient mood analysis from multimodal video," *25th IEEE International Conference on Image Processing (ICIP)*, 2018.

[15] D. T. Stuss and B. Levine, "Adult clinical neuropsychology: lessons from studies of the frontal lobes," *Annual review of psychology*, vol. 53, no. 1, pp. 401–433, 2002.

[16] H. Li and G. Hua, "Hierarchical-pep model for real-world face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4055–4064.

[17] J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua, "Neural aggregation network for video face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4362–4371.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[19] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1805–1812.

[20] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–10.

[21] M. Bellantonio, M. A. Haque, P. Rodriguez, K. Nasrollahi, T. Telve, S. Escalera, J. Gonzalez, T. B. Moeslund, P. Rasti, and G. Anbarjafari, *Spatio-temporal Pain Recognition in CNN-Based Super-Resolved Facial Images*. Cham: Springer International Publishing, 2017, pp. 151–162.

[22] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Information Sciences*, vol. 428, pp. 49–61, 2018.

[23] J. Wan, S. Escalera, G. Anbarjafari, H. Jair Escalante, X. Baró, I. Guyon, M. Madadi, J. Allik, J. Gorbova, C. Lin *et al.*, "Results and analysis of chalearn lap multi-modal isolated and continuous gesture recognition, and real versus fake expressed emotions challenges," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3189–3197.

[24] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 94–101.

[25] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics*, 2005, pp. 465–468.

[26] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2005, pp. 2785–2790.

[27] T. Salter, F. Michaud, D. Létourneau, D. C. Lee, and I. P. Werry, "Using proprioceptive sensors for categorizing human-robot interactions," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2007, pp. 105–112.

[28] M. Fujita, "On activating human communications with pet-type robot aibo," *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1804–1813, 2004.

[29] J. M. Kessens, M. A. Neerincx, R. Looije, M. Kroes, and G. Bloothooft, "Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1–7.

[30] C. Jayawardena, I. H. Kuo, E. Broadbent, and B. A. MacDonald, "Socially assistive robot healthbot: Design, implementation, and field trials," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1056–1067, 2016.

[31] S. Ueki, H. Kawasaki, S. Ito, Y. Nishimoto, M. Abe, T. Aoki, Y. Ishigure, T. Ojika, and T. Mouri, "Development of a hand-assist robot with multi-degrees-of-freedom for rehabilitation therapy," *IEEE/ASME Transactions on Mechatronics*, vol. 17, no. 1, pp. 136–146, 2012.

[32] T. Yokoo, M. Yamada, S. Sakaino, S. Abe, and T. Tsuji, "Development of a physical therapy robot for rehabilitation databases," in *2012 12th IEEE International Workshop on Advanced Motion Control (AMC)*, 2012, pp. 1–6.

[33] J. Saunders, D. S. Syrdal, K. L. Koay, N. Burke, and K. Dautenhahn, "x201c;teach me x2013;show me x201d; x2014;end-user personalization of a smart home and companion robot," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 27–40, 2016.

[34] K. Swift-Spong, E. Short, E. Wade, and M. J. Matarić, "Effects of comparative feedback from a socially assistive robot on self-efficacy in post-stroke rehabilitation," in *IEEE International Conference on Rehabilitation Robotics (ICORR)*, 2015, pp. 764–769.

[35] J. Fan, D. Bian, Z. Zheng, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A robotic coach architecture for elder care (rocare) based on multi-user engagement models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1153–1163, 2017.

[36] B. Goertzel, J. Mossbridge, E. Monroe, D. Hanson, and G. Yu, "Humanoid robots as agents of human consciousness expansion," *arXiv preprint arXiv:1709.07791*, 2017.

[37] L. Helding, "Howard gardner's theory of multiple intelligences," *Journal of Singing*, vol. 66, no. 2, p. 193, 2009.

[38] M. J. Johnson, M. A. Johnson, J. S. Sefcik, P. Z. Cacchione, C. Mucchiani, T. Lau, and M. Yim, "Task and design requirements for an affordable mobile service robot for elder care in an all-inclusive care for elders assisted-living setting," *International Journal of Social Robotics*, pp. 1–20, 2017.

[39] C. Datta, "Programming behaviour of personal service robots with application to healthcare," Ph.D. dissertation, ResearchSpace@ Auckland, 2014.

[40] A. Pandey and R. Gelin, "A mass-produced sociable humanoid robot: pepper: the first machine of its kind," *IEEE Robotics & Automation Magazine*, no. 99, pp. 1–1, 2018.

[41] F. Tanaka, K. Isshiki, F. Takahashi, M. Uekusa, R. Sei, and K. Hayashi, "Pepper learns together with children: Development of an educational application," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 270–275.

[42] Y. Wu, H. Liu, and H. Zha, "Modeling facial expression space for recognition," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005, pp. 1968–1973.

[43] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, second edition ed. Springer London Dordrecht Heidelberg New York: Springer, 2011.

[44] M. Wang and W. Deng, "Deep face recognition: A survey," *arXiv preprint arXiv:1804.06655*, 2018.

[45] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 7, pp. 711–720, 1997.

[46] W. Deng, J. Hu, J. Guo, H. Zhang, and C. Zhang, "Comments on" globally maximizing, locally minimizing: Unsupervised discriminant projection with application to face and palm biometrics"," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1503–1504, 2008.

[47] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 2037–2041, 2006.

[48] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.

[49] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1. IEEE, 2005, pp. 786–791.

[50] L. J. Karam and T. Zhu, "Quality labeled faces in the wild (qlfw): a database for studying face recognition in real-world environments," in *Human Vision and Electronic Imaging XX*, vol. 9394. International Society for Optics and Photonics, 2015, p. 93940B.

[51] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3025–3032.

[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[53] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.