

An Advanced Stress Detection Approach based on Processing Data from Wearable Wrist Devices

Mazin Alshamrani

The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research (HURI)
Umm Al-Qura University, Makkah, Saudi Arabia

Abstract—Today's busy lifestyle often leads to frequent stress, the accumulation of which may lead to severe consequences for humans. Smartwatches are widely distributed and accessible, and as such deserve intelligent solutions that deal with the processing of such collected data and ensuring the improvement of the quality of life of end-users. The goal of this research is to create a stress detection technology that can correctly, constantly, and unobtrusively monitor psychological stress in real time. Due to the importance of stress detection and prevention, many traditional and advanced techniques have been proposed likewise we provide a unique stress-detection technique that is context-based. Due to the importance of stress detection and prevention, many traditional and advanced techniques have been proposed. In this research, a novel approach to designing and using a deep neural network for stress detection is presented. To provide a desirable training environment for network development, an open-source data set based on motion and physiological information collected from wrist and chest-worn devices was acquired and exploited. Raw data were analyzed, filtered, and preprocessed to create the best possible training data. For the proposed solution to have wide use value, further focus was placed on the data recorded using only smartwatches. Smartwatches are widely distributed and accessible, and as such deserve intelligent solutions that deal with the processing of such collected data and ensuring the improvement of the quality of life of end-users. Finally, two network types with proven capabilities of processing time series data are examined in detail: a fully convolutional network (FCN) and a ResNet deep learning model. The FCN model showed better empirical performances, and further efforts were made to select an optimal network structure. In the end, the proposed solution demonstrated performance similar to state-of-the-art solutions and significantly better than some traditional machine learning techniques, providing a good foundation for reliable stress detection and further development efforts.

Keywords—Fully convolutional neural network; stress detection; smartwatch; data pre-processing; semi-supervised learning

I. INTRODUCTION

The definition of stress emphasizes that it is a natural phenomenon that occurs when an organism tries to adapt to a life problem, life challenge, event, or situation. In that sense, stress is any negative reaction of the organism that occurs due to an attempt to adapt the organism to some sudden or unpleasant influence. It is commonly manifested by mental or physical suffering. As one type of emotion, stress is, unfortunately, an almost everyday phenomenon in people's lives, and as such, it has always been a complex challenge for

its prevention, analysis, and monitoring. There are three main features that make measuring stress a difficult and worthwhile topic to investigate. The stress is quite subjective as it has a stimulus that initiates the stress response in one individual may not initiate it in another. In addition, the ground truth for stress detection is difficult to define because of the high subjectivity and ongoing nature of the stress process, defining the onset, length, and severity of a stress event is challenging. Furthermore, stress cannot be directly measured as its reaction is made up of physiological, behavioral, and emotional components [1, 2]. Therefore, wearable devices can directly measure a portion of the physiological reaction (e.g., increased heart rate, increased sweating rate, etc.). However, there are no direct ways to measure the other two components of the stress response (behavioral and emotional reaction). With the advancement of technology and many approaches to the treatment of stress data, artificial, and intelligent approaches for solving this issue have emerged. Some interesting research on different methodologies for treating emotions and even stress data can be seen in [1-5]. Using various previously listed papers and conducted research, hidden knowledge and unknown data patterns can be found. Even predictions of stressful events can be generated in an accurate manner.

The motivation for this new research is based on the wearable stress and affects detection (WESAD) data set from the University of California Irvine (UCI) machine learning repository, which was publicly introduced and presented for the first time in [6]. WESAD data, a collection of curated databases, are maintained by the UCI and freely available to the worldwide machine learning community. In [6] is included research that examines motion and physiological information acquired from the chest and wrist-worn devices while worn by 15 participants (12 male and three female participants), with an average age 27.5 years. Examined WESAD data includes three different affective states: neutral, stress, and amusement. Furthermore, in [6] is presented the classification linear discriminant analysis (LDA) model for processing data that achieved an F1-score accuracy of 91%. Another complete approach by the same author from [6] of identifying and labeling affective states is presented in [7]. Applicability of the WESAD data set is shown in a few other research papers, where the authors tried to achieve improved accuracy performances by exploiting different intelligent algorithms. For example, in [8], only wrist sensor measurements from the WESAD data set are exploited, highlighting that wrist data measuring techniques are non-intrusive and widely available for acquiring. The research [8] uses three different machine learning models (i.e., logistic regression, decision tree, and

random forest) without any previous feature engineering processes. The best performances were achieved with the random forest model, achieving an accuracy between 88% and 99%, depending on the exploited feature. The article [9] examines if stress can be reliably detected only by using sensor data from a smartwatch. For experimental purposes, the authors used only wrist WESAD data and demonstrated satisfactory stress detection accuracy by using three different models: LDA, quadratic discriminant analysis (QDA), and random forest. Once again, as in [6], LDA showed the best performances. The research of [9] is also valuable from the perspective that it provides insights into what combination of different sensors can provide the most useful data for stress detections, highlighting the next three measuring devices: heart rate (HR) sensors, blood volume pulse (BVP), and skin temperature (ST) sensors. The most recent state-of-the-art research for automated stress detection in real life [8], [9] suggest an approach that employs a chest sensor. In their method, they first fine-tune their machine-learning model in the lab before applying it in real-world situations with certain simplifications, such as excluding times of moderate to high activity. They propose smartwatches as a source of physiological data in the future, as well as improved handling of physical activity and adding context information in the process of stress detection. All of these concerns are addressed in this research effort by using a source of physiological data is a wrist gadget. This used to recognize the user's activity by utilizing a machine-learning technique to analyze the acceleration data from the wrist device. This model is using a real-world contextual information in the machine-learning process to increase the method's effectiveness. Furthermore, we study the problem of stress detection under laboratory settings first, using an off-the-shelf wrist gadget outfitted with bio-sensors, and then apply the derived laboratory knowledge to real life, using data obtained entirely in the wild. In addition to laboratory expertise, real-world context information is collected to ensure that the approach may be effectively applied to real-world data. The context information is necessary to distinguish between real-life psychological stress and the various circumstances that cause comparable physiological arousal (e.g., exercise, eating, hot weather, etc.).

Unlike in [6, 8, 9], in [10] are applied deep learning (DL) techniques are applied in [10] to provide desired results in processing the WESAD data set. The DL model is designed to possess the ability to receive data from network inputs with different sampling rates. For that purpose, four different classification sub-models are proposed, each processing a single input with a specified sample rate and making individual predictions on its output. Final classification values are calculated by applying the fusion mechanism and applying the random forest model to generate all sub-models' predictions. Fundamental information about the fusion mechanisms can be acquired in [11]. Another recent study based on DL techniques in processing the WESAD data is presented in [12], where self-supervised learning (SSL) methodology was used to augment the initial data. This paper is different from the others listed here because it used an additional three data sets in pair with the WESAD data set and exploited only the electrocardiogram (ECG) feature. The methodology in [12] includes two main learning steps: unsupervised and supervised. In the

unsupervised part of the model, its goal was to detect and recognize previously applied data transformations without introducing any pre-defined labels and creating the features. In the second part, a transfer learning approach from [13] was used for the supervised classification of affective states by using previously created features. Another study that is based on applying SSL techniques to the WESAD data set is presented in [14]. The paper uses a "pretext task" to train the model without using labeled data, where it must be determined whether the raw data and the wavelet transformations are temporally aligned. The proposed model in [14] is evaluated in two ways: using a linear classifier on top of the SSL component and assessing the number of used samples for the supervised learning process. The first evaluation approach includes a direct comparison of the features created by SSL with the features designed with expert knowledge, as in [6]. The second one is based on every participant's feedback, where they were individually asked to interactively provide input information when they field stress. The algorithm utilized this feedback information to classify stress for every subject of examination. This approach is possible only if an intelligent model does not require an extensive database with labeled data and can learn from very few provided labels. Another paper that proposes an efficient semi-supervised network architecture for classification purposes is presented in [15]. The highlighted advantages of the model are its good applicability to big data in medical diagnosis. One interesting fact of the model from [15] is that it can be applied in processing structured, semi-structured, and unstructured data at the same time.

The goal of the novel research in this paper is to propose an intelligent framework for stress detection by using only wrist data acquired from a smartwatch. By reviewing previously introduced papers [6, 8, 9, 10, 12], it can be concluded that the current stress detection methodologies suffer from two significant deficiencies: 1) common usage of highly intrusive ECG and electroencephalogram (EEG) sensors that are not available to a broad public and 2) difficulties in getting high quality and reliable data due to the complexity of reading affective state values from appropriate sensors. As in [9], this research will also seek to avoid intrusive sensors for collecting data and will focus only on wrist sensors available on commercial smartwatches. Additionally, good practices in work with SSL models in treating the WESAD data from [12, 14] will be used as a starting point for developing a novel SSL methodology. A review of exploited methods for collecting, processing, and evaluating data collected by wearables (smartwatches and bands) is presented in [16]. It provides useful insights into techniques for intelligent algorithms' practical applicability while operating with wearable sensing equipment. It is also shown that HR sensors, galvanic skin responses, and body temperature sensors should be of leading interest in collecting data when devices are restricted to smartwatches.

In [17], another modeling effort of the WESAD data set is presented that includes both feature engineering and DL techniques for processing the data. It was proven that the combination of multiple deep neural networks could provide high performances with an average of 97.2% recall and 97.7% precision within all examined classes. The proposed solution's

downside is the high complexity and significant computation costs caused by utilizing one separate network for processing data of one single sensor. Information from all the networks is finally concatenated, and final classifications are produced. One other DL application for stress and affect detection is presented in [18]. The approach utilizes recurrent neural networks and provides high-accuracy results, with 97.5% of accurately detected values.

Another approach to human stress level examination is presented in [19]. The influence of the urban environment causes stress, and the article is of interest for new research, bearing in mind that the data are collected only by using the wrist devices. Thirty people participated in the study, and the raw unlabeled data was recorded during the 30 h of the experiment. The data format is suitable for our future self-supervised training procedure, and this experiment will further be explained in following sections.

II. RESEARCH METHODOLOGY

The previous section presented the summation of different intelligent approaches in treating the multimodal WESAD data set. The main goal in all these approaches was data classification and predictions of the stress conditions of involved participants. Introduced state-of-the-art research was used as an initial foundation for designing a novel intelligent solution in the domain of stress detection and was presented in the following parts of this research paper.

To start with an in-depth analysis of the proposed solution, the examined data was presented first. Included WESAD features represented physiological and motion data recorded from both chest and wrist-worn devices. The following biological parameters were examined: BVP, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration. The data included expert features crafted by using widely established physiological knowledge and medical procedures that are mostly utilized to interpret respiration results and the heartbeat rate. Furthermore, the dataset contained information about three different affective states of participants (neutral, stress, and amusement), which represented the most critical parameters for this research. However, it should be highlighted that the stress conditions were restricted to public speaking exercises, and no other types of stress causes were analyzed. This is the primary deficiency of utilized data, considering that a model trained on this data might not perform at a desirable level on the general population. Besides the WESAD features, the previously introduced article [19] represents an essential base for additional data. It includes three different and associated open-source data sets that provide more than 50 h of raw Empatica E4 wrist measurements were used in this research for the semi-supervised learning phase. Furthermore, the specific Empatica E4 wrist device from [19] is of central importance for new research because it was used in our laboratory environment and in the original WESAD experiment. This implies that it was possible to combine or compare the measurement results from the described research with the measurement results in this study, ultimately leading to an accurate and reliable evaluation of the novel model. Besides building intelligent applications on the WESAD data

set, many independent attempts were made to analyze emotions and extract meaningful insights from collected data of emotional parameters [20–22]. In [20, 21], different techniques were used to recognize various emotions, understand these emotions, and understand the overall reasons for their occurrence. Additionally, in [22], the Deep Multi-Net CNN Model was used for violence recognition in video surveillance. In this paper, another emotional state that caused violent behavior was examined, but not by using internal human conditions and measurement of biological parameters, but by using recorded participants' video shots.

Finally, based on the previously introduced articles, it was concluded that there were few studies based on utilizing DL techniques on wrist data wearables. This research tried to fill the observed scientific gap and propose a new approach based on the combination of a DL algorithm with a semi-supervised learning mechanism. The methodology focused on the following four phases: data exploration and preparation, design and tuning of suitable DL models, application of prepared models on the optimized data set, and evaluation of performances and analysis of obtained results. In the next section, the first phase and applied exploratory data analysis are presented on the research data.

III. EXPLORATORY DATA ANALYSIS

Seventeen subjects (persons) participated in the original WESAD research, where they were labeled S1 to S17. This analysis is based on the WESAD data collection, which is freely available to the public. It comprises data collected from 17 individuals using the Empatica wrist-worn gadget. This gadget has accelerometers (ACC) as well as sensors for measuring skin temperature (ST), electrodermal activity (EDA), blood volume pulse (BVP), heart rate (HR), and heart rate variability (HRV). WESAD incorporates data from the chest-worn RespiBAN device, as well as questionnaires linked to participants' moods during the data collection session, in addition to E4 data. However, due to unreliable sensor results acquired in two cases, S1 and S12, these two specific subjects were removed from the research data in this paper. The rest of the data were used for building the required experimental data sets. Exploratory data analysis in this research was performed by using subjects S2 to S10 from the WESAD database as the features of the training set. The remaining subjects were assigned to the test and validation sets. With the purpose of preparing the data optimally for future DL processing, the initial data were treated in the following way. At first, the responses of all wrist sensors were aligned at the same timeline $f = 700\text{Hz}$. Moreover, all recorded sensor data from all included subjects were merged, and the overall data set was created accordingly. The data exploration phase was performed exclusively on the training data consisting of 40 million data rows. Keeping in mind that this research's main focus was stress detection, training data was initially analyzed from the perspective of the types of information within the set and the influence they could individually have on the stress feature. For the beginning, Fig. 1 graphically represents all subject activities that are registered during the measurement phase and saved to the training data set.

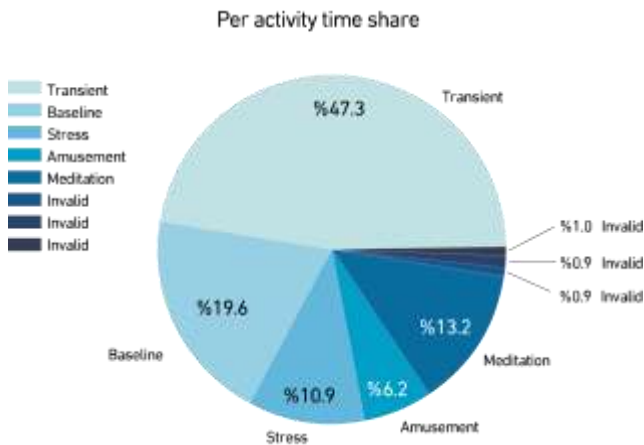


Fig. 1. The Share of Individual Activities within the Training Set.

It is easy to conclude by examining Fig. 1 that the utilized data set is imbalanced from the stress feature's perspective: Only 11% of training data is associated with stress occurrence. However, the quantity of the stress data expressed through time is 4 h, which should be sufficient material for a future model's training. Moreover, 3% of data are explicitly labeled by the data set authors as invalid and are to be removed from further work. Another question that should be answered concerning the research goal is which sensor and which recorded features have the most correlation to the stress status. For that purpose, correlation analysis was performed, and graphical results are presented in Fig. 2. It can be concluded from the figure that the acceleration and electrodermal activities (EDA) have the most correlation with the stress feature, recorded both on wrist and chest devices. This agrees with the intuitive and judgmental conclusions that stress generally causes an increase in breathing rate, chest acceleration, and sweating.

In the next phase of the data analysis and pre-processing, the outlier removal technique was performed. Each sensor was pre-defined with acceptable ranges of values, and measured values outside of these ranges were deleted and replaced by the closest valid values. Table I presents all needed information about exploited features and defined ranges.

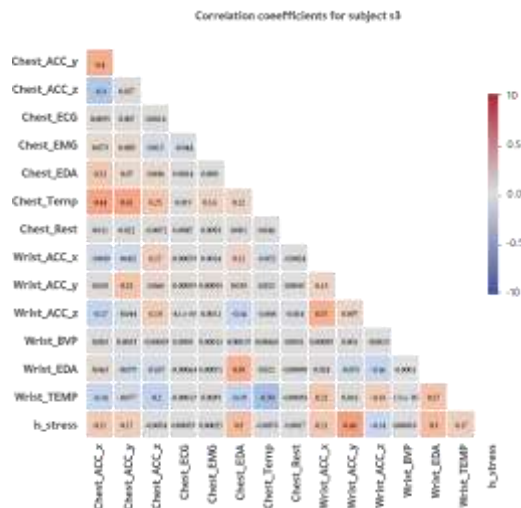


Fig. 2. Correlation Analysis of Training Data Set Concerning the Stress Feature.

TABLE I. EXAMINED SENSORS AND RELATED FREQUENCIES

Sensor	Sampling frequency (HZ)	Cutoff frequency	Defined range
Chest acceleration	700	70	X: [0;2] Y: [-0.5;0.5]. Z: [-2;2]
Chest ECG	700	70	[-1;1]
Chest EMG	700	70	[-0.5;0.5]
Chest EDA	700	2	[0;18]
Chest temperature	700	0.5	[30;36]
Chest respiration	700	2	[-10;10]
Wrist acceleration	32	16	X: [-100;150] Y: [-75;75] Z: [-75;100]
Wrist BVP	64	32	[-200;200]
Wrist EDA	4	2	[0;5]
Wrist temperature	4	2	[28;36]

Further, when working with environmental and real sensor measurements, the occurrence of noise is a common situation. Generally, any sensor signal is divided into two parts: a signal component that includes valuable information and a random noise component. In order to remove the noise component, a low-pass filter was utilized to remove the noise frequencies and undesirable data. The filtering procedure was performed as follows: a specific cutoff frequency was selected for each sensor, which represented the sensor's highest meaningful frequency values. The cutoff frequencies were selected by visual inspection of the signals, and their numerical values were provided in Table I. Next, a second-order Butterworth low-pass filter with the four corresponding cutoff information were utilized to process the signal. The example of successful filtering of one of the examined features is presented in Fig. 3.

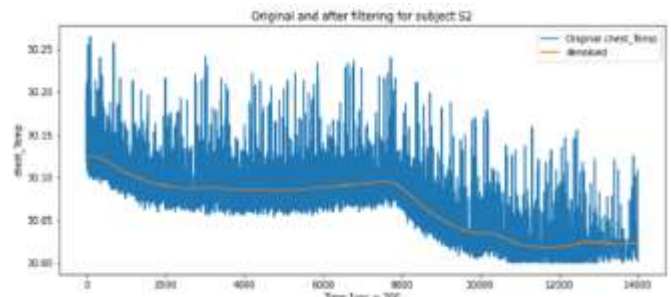


Fig. 3. An Example of Original and De-noised Signals of the Chest Temperature Feature.

IV. APPLICATION OF THE FULLY CONVOLUTIONAL NETWORK TO STRESS DETECTION TASKS

To design a model useful to a broad audience and applicable to almost any interested party, the training processes of neural network models focused only on using smartwatch WESAD data (wrist data). Following [9], all other sensor data from the initial data set that are not widely available in commercial smartwatches (like EDA sensors) were removed from the training data. For programming purposes, popular Python environments TensorFlow and Keras were used. Finally, a Google cloud machine with Nvidia K80 GPU was exploited to provide optimal computational power.

For research purposes, an FCN and a ResNet DL model from [23] were used for binary classification of baseline and stress states of the subjects. These models proved reliable and capable of a quality prediction of time series, especially the FCN model, a simple but effective model for time series classifications as shown in Fig. 4. Both networks were based on convolution layers, where the main difference between the two was in the number of layers: FCN was designed with 3, while ResNet possessed nine convolution layers. In general, a convolutional layer is a linear layer, like any other dense layer. However, the convolution layer structure was adapted for work with temporal information, which provided faster processing and improved accuracy of time series in comparison to a dense layer. In addition to these two types of networks, good results in DL analysis for emotion detection were presented [24].

After three convolutional layers of FCN, a global average pooling and the final SoftMax layer continued. At the end of each convolution, a batch normalization layer influencing training and convergence performances was applied. The main batch parameters, feature maps and striding, were tuned with special attention. Feature maps affected the total number of neurons within a network, while striding influenced how the network processed and sampled the time series data. The best empirical performances were achieved to combine four, two, and two strides for each of the three convolution layers of FCN. This combination was applied to the structure of the network. Feature maps layers 1 to 3 were selected by following the procedure from [23], and the following configuration was utilized: 64, 128, 256. The approach from [23] was also applied for selecting an Adam optimizer learning rate equal to 0.001, which was the default configuration for FCN. Two hundred and fifty learning epochs were specified, and the model that showed the best performances on the validation set was selected. Finally, ReLU activation functions [25] were chosen for building artificial neurons and DL models. Another approach to optimizing network parameters for a neural network-based emotion recognition framework was presented in [26].

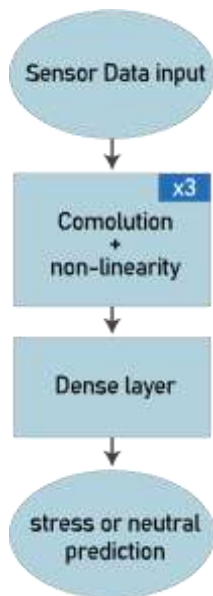


Fig. 4. Simplified Diagram of FCN Model.

For evaluation purposes, Leave-one-out cross-validation (LOO) from [27] was performed. The LOO experiment was performed on 15 folds, where for each tested subject, the data was trained on 12 other subjects, and two additional subjects were used for building the validation set. For example, if the test data was defined as S2, validation data variables were randomly selected to be S3 and S5, and all other variables were used for the training data set. By utilizing this kind of approach of treating acquired data, 15 different data sets were created through the experimental phase, and they allowed 15 different testing environments for the proposed intelligent algorithm, providing robust and reliable results in the end.

Different configurations of FCN and ResNet models were tested on the prepared training data, and it was observed that FCN was significantly faster (8 min) than ResNet (1h 30 min), and additionally, it performed better on processing the data (accuracy on examined sample: 81%–77%). The specific model configuration was selected by comparing achieved performances after a fixed number of epochs versus choosing the best model on the validation set. It was experimentally shown that the second approach provided more reliable performances, so the main testing parameter of the selection process was achieved on the validation set.

V. RESULTS ANALYSIS

In Fig. 5, the model performances for each examined subject are presented. An in-depth review of the achieved performances of the proposed model is presented in Table II. Metrics within the table can be explained as follows: accuracy—the number of correct classifications over total samples; balanced accuracy—the average of the proportion corrects for each class individually; F1—a harmonic average of precision and recall for the “stress” class; WEIGHTED-F1—similar to F1, this is an averaging of the “stress” and the “non-stress” class; area under curve (AUC)—a classification metric not impacted by class imbalance; precision—true “stress” detection overall stress detections; recall—true “stress” detections of overall stress samples.

Summarized results from Table II and the overall classification results of this research are presented in Table III. The average accuracy of the proposed model is approximately 0.85, while for the same training conditions, a conventional naive classifier achieved an accuracy of 0.78. Achieved results compares with the results from [8], where smartwatch data was also used. It is shown that the model from [8] provided slightly better accuracy from the model in this research (0.874 in comparison to 0.85). On the other hand, they performed 255 different runs during the training process of their model compared to only 15 performed runs in this research. It can be concluded that the proposed model in this research demonstrated satisfactory prediction performances with a small number of training cases. Furthermore, it should be reliably assumed that the model will likely further improve performance by providing additional training data and test cases and making additional tuning attempts.

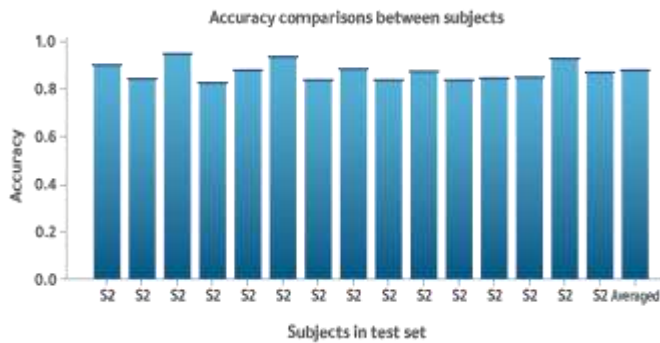


Fig. 5. Model Performances per each Examined Subject.

TABLE II. ACCURACY PER SUBJECT

Subject	Accuracy	Balanced Accuracy	F1	Weighted F1	AUC	Precision	Recall
S2	0.87	0.85	0.73	0.88	0.92	0.67	0.8
S3	0.81	0.67	0.5	0.8	0.86	0.61	0.43
S4	0.92	0.82	0.78	0.91	0.99	0.96	0.65
S5	0.8	0.87	0.68	0.82	0.98	0.52	0.99
S6	0.85	0.75	0.63	0.85	0.89	0.69	0.57
S7	0.91	0.91	0.81	0.91	0.96	0.72	0.91
S8	0.81	0.66	0.47	0.79	0.8	0.64	0.37
S9	0.86	0.7	0.56	0.84	0.82	0.82	0.43
S10	0.81	0.66	0.49	0.79	0.85	0.69	0.38
S11	0.85	0.78	0.67	0.85	0.9	0.67	0.67
S13	0.81	0.69	0.53	0.8	0.87	0.58	0.48
S14	0.82	0.63	0.4	0.79	0.87	0.82	0.27
S15	0.82	0.71	0.57	0.82	0.8	0.64	0.51
S16	0.9	0.92	0.81	0.9	0.97	0.7	0.96
S17	0.84	0.73	0.62	0.83	0.87	0.76	0.52
Averaged	0.85	0.76	0.62	0.84	0.89	0.7	0.6

TABLE III. IN-DEPTH ANALYSIS OF ACHIEVED RESULTS

Metric	FCN
Accuracy	0.85
Recall "baseline"	0.92
Precision "baseline"	0.89
F1 score "baseline"	0.90
Recall "stress"	0.59
Precision "stress"	0.68
F1 score "stress"	0.63

VI. CONCLUSION

In this research, a deep convolutional neural network model for stress detection was proposed. The model was implemented by exploiting only commercial smartwatch data because of the desire to provide a broad audience with a universal intelligent solution. Stress is a major issue in

today's society, with both social and economic consequences. The results show that these works have a high accuracy for identifying stress. However, because stress levels in everyday life can differ considerably from stress levels generated in laboratory conditions, daily life studies have gained popularity in the scientific community. Another important reason why everyday life stress detection studies are more appealing to researchers is because consumers do not desire intrusive measuring techniques employed in laboratory settings. Inconspicuous wearable gadgets can be used to assess stress levels in everyday life without disturbing the users. We covered open research issues for everyday life stress detection in this part, and there is still space for development in this area. Furthermore, the accuracies of stress detection methods in everyday life are significantly lower than those in laboratory conditions. The ultimate goal of stress detection is to create a high-accuracy scheme in everyday life by conquering unsolved difficulties and employing emotion management strategies to reduce the users' stress.

REFERENCES

- [1] Lisetti, C. L., and Nasoz, F. (2004) Using non-invasive wearable computers to recognize human emotions from physiological signals. EURASIP journal on applied signal processing, 2004, 1672–1687.
- [2] Pollreis, D., and Taheri, N. (2017) A simple algorithm for emotion recognition, using physiological signals of a smart watch. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11-15 July, 2353–2356, IEEE, New York, USA.
- [3] Albraikan, A., Hafidh, B., and Saddik, A. El. (2018) iaware: A real-time emotional biofeedback system based on physiological signals. IEEE Access, 6, 78 780–78 789.
- [4] Ragot, M., Martin, N., Em, S., Pallamin, N., and Diverrez, J. M. (2017) Emotion recognition using physiological signals: laboratory vs. wearable sensors. In International Conference on Applied Human Factors and Ergonomics, Los Angeles, California, 17- 21 July, 15-22, Springer, Berlin, Germany.
- [5] Zhao, Z., Wang, Z. Y., and Guo, B. (2018) Emotionsense: Emotion recognition based on wearable wristband. In 2018 IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (Smart-World/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), China, 8-12 October, 346-355, IEEE, New York, USA.
- [6] Schmidt, P., Reiss A., Duerichen R., Marberger, C., and Laerhoven, K. V. (2018) Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. ICMI '18: Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder CO USA, 16-20 October, 400-408, ACM, New York, USA.
- [7] Schmidt, P., Reiss, A., Dürichen, R., and Laerhoven, K. V. (2018) Labelling affective states in the wild: Practical guidelines and lessons learned. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 9-11 October, 654–659, Association for Computing Machinery, New York, United States.
- [8] Indikawati, F. I., and Winiarti, S. (2019). Stress Detection from Multimodal Wearable Sensor Data. IOP Conference Series: Materials Science and Engineering, Yogyakarta, Indonesia, 16 November, 1-6, IOP Publishing Ltd, Bristol, United Kingdom.
- [9] Siirtola, P. (2019) Continuous stress detection using the sensors of commercial smartwatch. UbiComp/ISWC '19 Adjunct: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, London, United Kingdom, 11-13 September, 1198–1201, Association for Computing Machinery, New York, NY, United States

- [10] Lin, J., Pan, S., Lee, C., and Oviatt, S. L. (2019) An Explainable Deep Fusion Network for Affect Recognition Using Physiological Signals. *CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, 3-7 November, 2069-2072, Association for Computing Machinery, New York, United States.
- [11] Popescu, A., Mocanu, I., and Cramariuc, B. (2020) Fusion Mechanisms for Human Activity Recognition Using Automated Machine Learning. *IEEE Access*, 8, 143996-144014.
- [12] Sarkar, P., and Etemad, A. (2020) Self-supervised ECG Representation Learning for Emotion Recognition. *IEEE Transactions of Affective Computing*, early access, accepted for publication.
- [13] Weiss, K., and Khoshgoftaar, T. (2017) Evaluation of Transfer Learning Algorithms Using Different Base Learners. *IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, Boston, USA, 6-8 November, 187-196, IEEE, New York, USA.
- [14] Saeed, A., Salim, F. D., Ozcelebi, T., and Lukkien, J. (2020) Federated Self-Supervised Learning of Multi-Sensor Representations for Embedded Intelligence. *IEEE Internet of Things Journal*, early access, accepted for publication.
- [15] Wang, L., Qian, Q., Zhang, Q., Wang, J., Cheng, W., Yan, W. (2020) Classification Model on Big Data in Medical Diagnosis Based on Semi-Supervised Learning, *The Computer Journal*, bxaa006.
- [16] Saganowski, S., Dutkowiak, A., Dziadek, A., Dzieżyc, M., Komoszyńska, J., Michalska, W., Polak, A., Ujma, M., and Kazienko, P. (2020) Emotion Recognition Using Wearables: A Systematic Literature Review. *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Austin, TX, USA, 23-27 March, 1-6, IEEE, New York, USA.
- [17] Chakraborty, S., Aich, S., Joo, M., Sain, M., and Kim, H. C. (2019) A Multichannel Convolutional Neural Network Architecture for the Detection of the State of Mind Using Physiological Signals from Wearable Devices. *Journal of Healthcare Engineering*, 2019, 1-17.
- [18] Ragav, A., Krishna, N. H., Narayanan, N., Thelly, K., and Vijayaraghavan, V. (2019). Scalable Deep Learning for Stress and Affect Detection on Resource-Constrained Devices. *18th IEEE International Conference on Machine Learning And Applications (ICMLA)*, Boca Raton, Florida, USA, 16-19 December, 1585-1592, IEEE, New York, USA.
- [19] Ojha, V. K. et al. (2019) Machine Learning Approaches to Understand the Influence of Urban Environments on Human's Physiological Response. *Information Sciences*, 474, 154-169.
- [20] Ratre, A. (2020) Stochastic Gradient Descent–Whale Optimization Algorithm-Based Deep Convolutional Neural Network To Crowd Emotion Understanding, *The Computer Journal*, 63(2), 267–282.
- [21] Afdhal, R., Ejwali, R., and Zaied, M. (2020) Emotion Recognition by a Hybrid System Based on the Features of Distances and the Shapes of the Wrinkles, *The Computer Journal*, 63(3), 351–363.
- [22] Mumtaz, A., Sargano, A. B., and Habib, Z. (2020) Fast Learning Through Deep Multi-Net CNN Model For Violence Recognition In Video Surveillance, *The Computer Journal*, bxaa061.
- [23] Wang, Z., Yan, W., and Oates, T. (2017) Time series classification from scratch with deep neural networks: A strong baseline. *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, Alaska, USA, 14-19 May, 1578-1585, IEEE, Piscataway, New Jersey, USA.
- [24] Kanjo, E., Younis, E. M., and Ang, C. S. (2019) Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, 49(-), 46–56.
- [25] Glorot, X., Bordes, A. and Bengio, Y. (2010) Deep Sparse Rectifier Neural Networks. *Journal of Machine Learning Research*, 15, 315-323.
- [26] Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F. and Chandran, V. (2018) Long short-term memory hyperparameter optimization for a neural network-based emotion recognition framework. *IEEE Access*, 6, 49 325–49 338.
- [27] Vehtari, A., Gelman, A. and Gabry, J (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27, 1413–1432.