# WEARABLE STRESS AND AFFECT DETECTION USING DEEP LEARNING

1st Ege Sendogan
*Department of Computer Engineering*
*Galatasaray University*
Istanbul, Turkey
https://orcid.org/0000-0002-4808-347X

***Abstract*—Automatic detection of stress and affective states is becoming increasingly popular in the human-computer interaction field. Besides that, it helps for monitoring and preventing mental health disorders. Thus, there's a lot of study in this field. However, the most of them employs classical machine learning methods. In this work, a new deep learning model for affect and stress detection is presented. The model trained and tested with WESAD dataset that contains wearable sensor data for stress and affect detection. At the end of the experiments, the average of the accuracy scores obtained was equal to %81.**

## I. Introduction

Emotional computing is a topic that has attracted the attention of researchers recently. There are several advantages of being able to automatically detect people's stress levels and moods.

Although there's been a great progress in human-computer interaction studies, the goal of making machines understand and analyse human emotions is still not achieved [1]. This deficiency in the field of human-computer interaction, where a lot of research made in recent years, will be eliminated when successful emotional state detection systems are developed.

On the other hand, stress is a natural reaction of the body to any danger that requires adaptation or response. The observed situation related to danger can be a real event or a situation that the mind perceives as "dangerous". It has been proven by long years of research that long-term stress has many damages on human health [12]. According to these studies, stress can lead to gastrointestinal problems [13], skin [14], heart [15], and even brain diseases [16] and so on. Considering these, it is vital to monitor and detect stress.

During stress, some biochemical changes occur in the body. These are changes, such as increased blood sugar, increased blood pressure, increased heartbeat, increased body temperature and sweating. These changes can be monitored by recording the electrical activity of the heart and brain, muscle activity, blood pressure, sweating level, breathing rate and body temperature with sensors such as ECG, EEG, EMG, BVP, EDA, EOG, RESP, and TEMP.

In recent years, the data obtained with aforementioned sensors have been used to train some machine learning and deep learning models to classify affective states. In previous years, machine learning models were preferred rather than deep learning models for classifying affective states. One of the main reasons for this is that deep learning models are quite resource demanding and therefore cannot be used in wearable devices which have relatively low computational capabilities.

On the other hand, in order to train machine learning models with sensor data, it is necessary to apply a feature extraction process on the data before the training phase. This makes it necessary to work with experts in this field to make sense of the data received from the sensors. As a consequence, there is a lack of affect detection dataset as well as there is a limited number of studies conducted to develop a system that detects affective state and stress.

However, very successful models have been developed in the field of deep learning in recent years, and the performance of deep learning models has surpassed the performance of machine learning models. Simultaneously, the computing capabilities of wearable devices have also improved.

In addition, technologies, such as fog computing and edge computing where these devices can offload computing loads have also emerged. As a result, the use of deep learning models for affective state classification has become more popular. This helps eliminating the feature extraction steps, as deep learning models learn the features themselves. In other words, there is no need for feature extraction anymore, which requires expert knowledge.

## II. Related Works

### A. Classical Methods

Classical methods have been used in most of the studies up to now for the classification of affective states. Classical methods typically consist of sequential stages of data preprocessing, signal transformation, feature extraction, feature extraction and use of classical machine learning algorithms such as k-nearest neighbors, support vector machines and decision trees.

In 2016, Grojeski et al. fed machine learning models with 63 features such as mean, standard deviation, quartiles, and quartile deviation, which they extracted from the data set they created. They preferred decision trees for the stress classification method, and obtained an accuracy of 72%.

In 2018, Subramanian et al. extracted some statistics, such as mean, standard deviation (std), skewness, kurtosis of the raw feature over time from ECG data and fed SVM and Naive

Bayes classifiers with them which lead them to an accuracy of 60%.

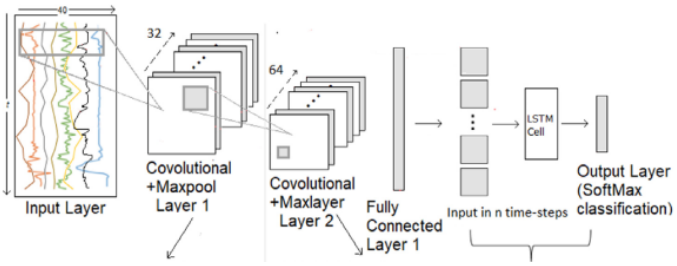


Fig. 1. CNN-LSTM Architecture

Detection of stress, emotions and mental illnesses is one of the important topics of human computer interaction. Therefore, it has attracted the attention of researchers for a long time.

Classical machine learning methods have been applied in many studies such as [10] and [11]. Although successful results have been obtained, there are some disadvantages of these methods. Because, in order for these methods to succeed, it is very valuable to apply a highly successful feature extraction process on the data. Due to the fact that this process includes high-level understanding of signals and sensors, it can be quite troublesome sometimes [1]. As a result, it can prevent many researchers from doing studies in this field.

### *B. Deep Learning Methods*

Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher-level features from the raw input. One of the biggest benefits of deep neural networks is their ability to perform automatic feature extraction from raw data, also called feature learning.

Despite this advantage, because of the fact that the employment of deep artificial neural networks is a quite fresh idea they are less preferred in the solution of the affective state classification problem. Therefore, there is not much source code that can be used in this field, and there are few other studies to compare models. An example of deep learning models is given in figure 1.

In 2019, E. Kanjo et al. presented two different architectures. The first is a convolutional neural networks based on an architecture consisting of 2 convolutional layers, 2 max pooling layers, and a fully connected layer. When this model was trained with data obtained with body sensors, a result of 0.709 F-measure was obtained. The other is the CNN-LSTM architecture that they created by adding a long short term memory layer to the end of the first model, and a result of 0.874 F-measure was obtained.

In 2020, Dzieżyc et al. tried 10 different architectures to create a benchmark for affective state detection in their study. Among these architectures, they reported that the most successful ones are Stresnet (the spectrotemporal residual network), FCN (the fully convolutional neural network) and Resnet (the residual network). FCN consists of parallel layers containing three convolutional layers and a common fully

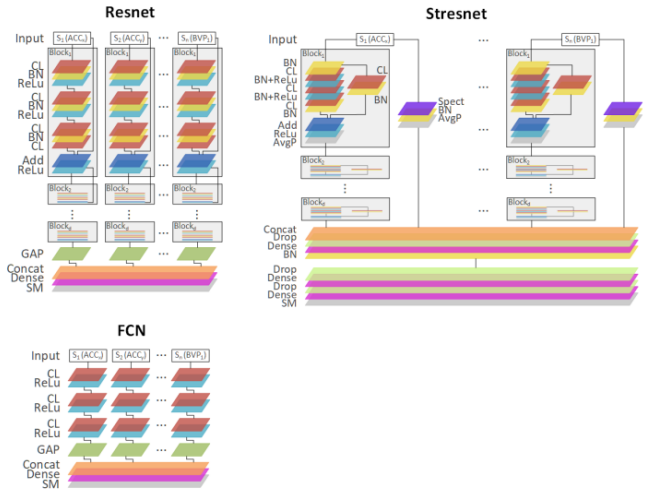


Fig. 2. FCN, Resnet, Stresnet

connected layer that takes the outputs of these parallel layers as an input and produces results. There is a similar "parallel branch" structure in Resnet. But in this architecture, each parallel branch has three blocks and each block has three convolutional layers, three batch norm layers and three relu layers. Stresnet's structure is similar to Resnet. The main difference between them is that while 1 signal is processed in each branch in Resnet, 2 signals are processed in each branch in Stresnet.

Another approach for affective state classification task is to use bidirectional LSTM and GRU based networks. For example, Li at al. employed an attention-based bidirectional LSTM network which uses spectrograms that are obtained by signal data as inputs and got an F1-score of 0.72. Similarly, Qiu at al. preferred an attention-based bidirectional GRU network and got satisfying results.

## III. Methodology

### *A. Proposed Methods*

The study of Dzieżyc et al. is very important because with the various deep learning architectures they implemented in this study, they completed a significant lack of benchmarks in deep learning-based affective state detection systems.

In their study, they adapted several deep learning architectures for time-series classification which are presented in [8]. Among those architectures, the FCN (the fully convolutional neural network) a model which contains three consecutive convolutional layers for each different signal channel in data set and a fully connected layer which gathers these different channels, has reached the highest accuracy and F1 scores on WESAD dataset [9] and proved the power of convolutional neural networks in the solution of affective state detection problem. On the other hand, the Resnet model that have several residual blocks consist of three convolutional layers similar to the FCN model, has reached one of the best accuracy and F1 scores, showed that a multi-tier architecture can be a potential architecture that can be used in similar projects.

TABLE I
EXISTING ARCHITECTURES AND THE ARCHITECTURE INTRODUCED IN THIS ARTICLE

| Architecture | Design |
|---|---|
| FCN | 14x[CL-CL-CL]-FC |
| Time-CNN | 14x[CL-CL]-FC |
| *Time-FCN* | 14x[[FCN]-[Time-CNN]]-FC |

Besides that, Time-CNN was another model that performs well on WESAD dataset in the study of Dzieżyc et al. The difference of this model is that it has a simpler architecture than the others. Briefly it gets its strength from its simplicity.

Based on the aforementioned inferences, I designed a new architecture that combines the power of convolutional neural networks, multi-tier and simple architectures for the solution of stress and affective state detection problem.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K$$

Fig. 3. Softmax Function

In this architecture, I used a FCN block and a Time-CNN block that takes the output of FCN block as inputs. Besides, Time-CNN block consists of two 1D convolutional layers that create a convolution kernel that is convolved with the layer input over a single spatial (or temporal) dimension to produce a tensor of outputs, and two average pooling layers that calculate the average value for each patch on the feature map.

The last layer of the T-CNN block is a flatten layer that transform the entire pooled feature map matrix into a single column. The outputs of second layer in each signal channels are concatenated and fed into a dense layer with sotmax activation function which normalizes the output of the network to a probability distribution over predicted output classes. This model is implemented in python language using tensorflow and keras frameworks.

### *B. Datasets*

In this project, WESAD data set is used [9]. WESAD is a multimodal data set created with pyhsical and motion data (ACC, ECG, BVP, EDA, EMG, RESP and TEMP) collected from 15 participants with wrist-worn (Empatica E4) and a chest-worn (RespiBAN) devices. It contains 3 different affective states (neutral, stress and amusement) and is frequently used in recent studies. The original sampling frequency of RespiBAN device for each modalities is 700Hz while the original sampling frequencies of Empetica device for BVP, ACC, EDA and TEMP signal modes are 64Hz, 32Hz, 4Hz, 4Hz.

Although we eliminated the feature extraction process by the virtue of deep learning, we need to apply a preprocessing process to achieve successful results. In this project, following preprocessing steps which are defined in [1]:

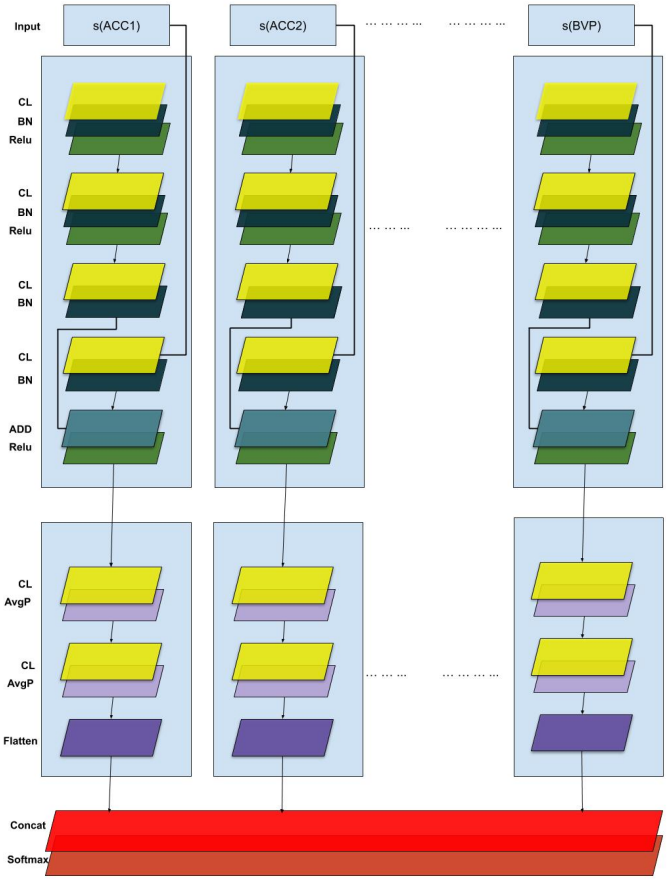


Fig. 4. T-CNN Model

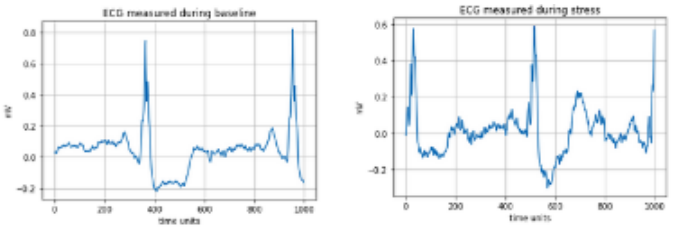


Fig. 5. An Example Data Graphic From WESAD Dataset

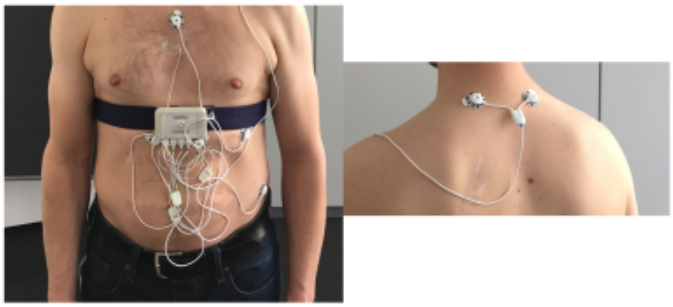

Fig. 6. Placement of the sensors

- Removal of outliers from the data
- Butterworth low-pass filter with a 10 Hz cut-off
- Downsampling. The downsampled frequencies of RespiBAN device for ECG, ACC, EMG, EDA, TEMP and Respiration signal modes are 70Hz, 10Hz, 10Hz, 3.5Hz, 3.5Hz, 3.5Hz
- Min-max normalization
- Windowing. WESAD dataset contains some large signals which are difficult to analyze statistically. In order to avoid this problem, each signal is divided into 60s windows with 30s slides.

Preprocessing steps are implemented in python language and several python libraries such as numpy, scipy, pandas, sklearn and math are used.

## IV. Experiments

All convolutional layers in the fully convolutional network block has 64 filters, with kernel sizes 8, 5, and 3. In addition, every convolutional layer is followed by a batch normalization layer as well as an activation layer with the relu (rectified linear units) function which is one of the most commonly used activation functions in deep learning models.

Besides, the first convolutional layer in the time cnn block has 6 filters with a kernel size of 7 while the second convolutional layer has 12 filters with a kernel size of 7. The rectified linear units is also used in this block as the activation function. This activation function is known for its ability of facilitating gradient descent and increasing training speed.

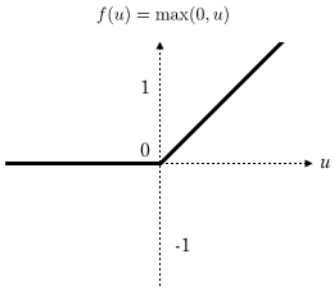


Fig. 7. ReLU function

Cross-Validation is an essential tool in the Data Scientist toolbox. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. Thus, a 5-fold cross validation is used in this project.

TABLE II
Design of Experiments

| | Training Sets | Test Sets | Validation Sets |
|---|---|---|---|
| Exp.1 | 5, 16, 15, 8, 9, 2, 7, 14, 6, 11 | 4, 3, 13 | 10, 17 |
| Exp.2 | 4, 16, 15, 17, 3, 9, 10, 14, 6, 11 | 5, 8, 7 | 2, 13 |
| Exp.3 | 4, 15, 17, 3, 8, 2, 10, 13, 6, 11 | 16, 9, 14 | 5, 7 |
| Exp.4 | 5, 16, 3, 8, 9, 10, 13, 7, 14, 11 | 15, 2, 6 | 17, 4 |
| Exp.5 | 4, 5, 16, 3, 8, 9, 2, 13, 7, 14 | 17, 10, 11 | 6, 15 |

As a result, five different experimental setups with training, test and validation data are created from the WESAD data set which is created with data from 15 different subjects. In each experimental setup, test data is created with the 1/5 of the subject's data.

The reason that a validation set also created apart from the test set in every setup is to prevent overfitting. Besides that, training lasted 1000 epochs with a batch size of 16 in each experiment.

After 1000 epochs in each experimental setup, the weights of the model with the lowest validation loss were recorded, and then the model with these weights was tested with the test data of the experimental setup.

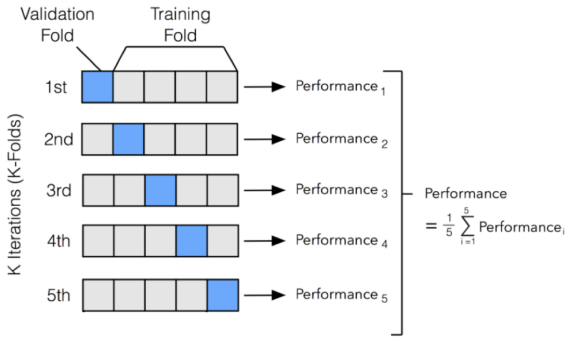


Fig. 8. 5-fold cross validation

## V. Results

The classification task in this project was to classify the affective state in each time unit with the BASELINE, AMUSEMENT and STRESS classes, using signal data from wearable sensors.

The accuracy and macro F1-score are utilized in order to compare results. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. Macro F1-score will give the same importance to each label/class. It does not take label imbalance into account.

Table 2 shows the results of the experiments. The average accuracy rate is equal to 0.81 while average F1-score is equal

$$F1\ score\ macro = \frac{F1\ score_1 + \ldots + F1\ score_N}{N}$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Fig. 9. The metrics

TABLE III
RESULTS OF EXPERIMENTS

| | Precision | Recall | Accuracy | F1-score |
|---|---|---|---|---|
| Exp. 1 | 0.88 | 0.66 | 0.81 | 0.63 |
| Exp. 2 | 0.74 | 0.75 | 0.79 | 0.74 |
| Exp. 3 | 0.85 | 0.88 | 0.89 | 0.86 |
| Exp. 4 | 0.66 | 0.65 | 0.76 | 0.65 |
| Exp. 5 | 0.78 | 0.73 | 0.79 | 0.75 |
| Avg | 0.78 | 0.73 | 0.81 | 0.73 |

to 0.73 . While the accuracy rate obtained in the experiments is close to each other except for the third experiment, F1 scores range between 0.63 and 0.86. The highest accuracy rate was obtained in the third experiment and is equal to 0.89. Also, the lowest accuracy achieved in experiments is equal to 0.76.

As mentioned before, there are quite a few studies using deep learning in researches related to affective state and stress detection. In addition to this, finding an article using the WESAD data set with deep learning-based systems is even more difficult. Fortunately, Dzieżyc et al. created a benchmark with the study where they presented various deep learning architectures for affect recognition from physiological sensor data sets including WESAD.

In their work, they created a total of 12 different architectures, including FCN and Time-CNN, which were the inspiration for the Time-FCN architecture presented in this article. Besides, they trained and tested these architectures with AMIGOS, DECAF, ASCERTAIN and WESAD datasets. The results that they obtained with WESAD dataset is presented in figure 10.

As seen, FCN and Time-CNN models are the models with the highest accuracy. In addition, the FCN model is the model with the highest F1-score, while the Time-CNN model is the model with the third best F1-score.

Besides that, it can be said that the Time-FCN model that I presented in this project is also very successful, and it is one of the models that can be used in affective state detection.

## VI. CONCLUSION AND FUTURE WORKS

The aim of this project is, by using deep learning, to match the data obtained from the human body with wearable sensors with three classes (baseline, amusement, stress) that depict people's moods .

| Architecture | Accuracy | ± | F1-Score | ± | ROC AUC | ± |
|---|---|---|---|---|---|---|
| FCN | 0.79 | 0.05 | 0.73 | 0.07 | 0.91 | 0.02 |
| Resnet | 0.74 | 0.07 | 0.69 | 0.07 | 0.89 | 0.04 |
| Time-CNN | 0.75 | 0.03 | 0.66 | 0.05 | 0.86 | 0.02 |
| MCDCNN | 0.74 | 0.04 | 0.62 | 0.05 | 0.84 | 0.03 |
| Stresnet | 0.69 | 0.11 | 0.62 | 0.10 | 0.82 | 0.05 |
| MLP-LSTM | 0.73 | 0.01 | 0.61 | 0.03 | 0.82 | 0.01 |
| Inception | 0.71 | 0.06 | 0.58 | 0.07 | 0.81 | 0.07 |
| Encoder | 0.71 | 0.03 | 0.57 | 0.05 | 0.83 | 0.02 |
| MLP | 0.72 | 0.01 | 0.57 | 0.01 | 0.78 | 0.02 |
| CNN-LSTM | 0.70 | 0.02 | 0.56 | 0.03 | 0.79 | 0.02 |
| Random guess | 0.33 | – | 0.32 | – | – | – |
| Majority class | 0.53 | – | 0.23 | – | – | – |

Fig. 10. Results obtained in [1]

For this purpose, WESAD dataset which is a publicly available dataset for wearable stress and affect detection is used. This dataset includes different modalities such as blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration as well as it contains three different affective states such as neutral, stress, amusement. Thus, this data set is the most suitable public data set for the purpose of our project.

Besides, WESAD dataset is a time series dataset. Time series data is a collection of observations obtained through repeated measurements over time.

The power of convolutional neural networks in classifying time series has been proven in previous studies and various models have been designed so far. ResNet architecture that consist of multiple convolutional neural network blocks and FCN (Fully Connected Network) which takes its power from the simplicity of its architecture, are examples of convolutional neural network based architectures for classifying time series.

In this project, a new convolutional neural networks based deep learning architecture for times series classification is proposed. This architecture contains two different consecutive CNN blocks. It has a very simple architecture and it gets its strength from this simplicity. Five different experiments were run on the designed model and quite satisfactory results were obtained.

Although CNN based models are examined in this project, some RNN (recurrent neural networks) based models were also used in deep learning for affective state and stress detection projects. This project can be extended by examining the performance of RNN based models as well as models that contain both recurrent and convolutional layers on the same data set. In addition, we can augment the training data by collecting data in addition to WESAD dataset and so we can train our models with larger datasets.

## REFERENCES


[1] Dzieżyc, M., Gjoreski, M., Kazienko, P., Saganowski, S., Gams, M. (2020). Can We Ditch Feature Engineering? End-to-End Deep Learning for Affect Recognition from Physiological Sensor Data. Sensors, 20(22), 6535. doi:10.3390/s20226535

[2] M. Gjoreski, H. Gjoreski, and M. Gams. 2016. Continuous stress detection using a wrist device: In laboratory and real life. In UbiComp '16. 1185–1193.

[3] J. Kim and E. André. 2008. Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 12 (2008), 2067–2083

[4] Kanjo, E.; Younis, E.M.; Ang, C.S. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. Inf. Fusion 2019, 49, 46–56.

[5] Subramanian, R.; Wache, J.; Abadi, M.K.; Vieriu, R.L.; Winkler, S.; Sebe, N. ASCERTAIN: Emotion and personality recognition using commercial sensors. IEEE Trans. Affect. Comput. 2016, 9, 147–160.

[6] Sarkar, P., Etemad, A. (2020). Self-supervised ECG Representation Learning for Emotion Recognition. IEEE Transactions on Affective Computing, 1-1. doi:10.1109/taffc.2020.3014842

[7] Saganowski, S.; Dutkowiak, A.; Dziadek, A.; Dzieżyc, M.; Komoszyńska, J.; Michalska, W.; Polak, A.; Ujma, M.; Kazienko, P. Emotion Recognition Using Wearables: A Systematic Literature Review—Work-in-progress.

[8] Fawaz, H.I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A review. Data Min. Knowl. Discov. 2019, 33, 917–963.

[9] Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder, CO, USA, 16–20 October 2018; pp. 400–408.

[10] Iacoviello D, Petracca A, Spezialetti M, Placidi G. A real-time classification algorithm for EEG-based BCI driven by self-induced emotions. Comput Methods Programs Biomed. 2015 Dec;122(3):293-303. doi: 10.1016/j.cmpb.2015.08.011. Epub 2015 Aug 29. PMID: 26358282.

[11] Khezri M, Firoozabadi M, Sharafat AR. Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals. Comput Methods Programs Biomed. 2015 Nov;122(2):149-64. doi: 10.1016/j.cmpb.2015.07.006. Epub 2015 Jul 29. PMID: 26253158.

[12] Mariotti A. (2015). The effects of chronic stress on health: new insights into the molecular mechanisms of brain-body communication. Future science OA, 1(3), FSO23. https://doi.org/10.4155/fso.15.21

[13] Huerta-Franco, M. R., Vargas-Luna, M., Tienda, P., Delgadillo-Holtfort, I., Balleza-Ordaz, M., Flores-Hernandez, C. (2013). Effects of occupational stress on the gastrointestinal tract. World journal of gastrointestinal pathophysiology, 4(4), 108–118. https://doi.org/10.4291/wjgp.v4.i4.108

[14] Kimyai-Asadi, A., Usman, A. (2001). The Role of Psychological Stress in Skin Disease. Journal of Cutaneous Medicine and Surgery, 5(2), 140–145. https://doi.org/10.1177/120347540100500208

[15] Chandola, T., Britton, A., Brunner, E., Hemingway, H., Malik, M., Kumari, M., ... Marmot, M. (2008). Work stress and coronary heart disease: what are the mechanisms?. European heart journal, 29(5), 640-648.

[16] Jung, S., Choe, S., Woo, H., Jeong, H., An, H. K., Moon, H., Ryu, H. Y., Yeo, B. K., Lee, Y. W., Choi, H., Mun, J. Y., Sun, W., Choe, H. K., Kim, E. K., Yu, S. W. (2020). Autophagic death of neural stem cells mediates chronic stress-induced decline of adult hippocampal neurogenesis and cognitive deficits. Autophagy, 16(3), 512–530. https://doi.org/10.1080/15548627.2019.1630222