Comparison of Stress Detectors based on Physiological Signals

Angela Mayhua Quispe
PhD Student in Computer Science
Universidad Nacional de San Agustin
Arequipa, Peru
amayhuaq@unsa.edu.pe

***Abstract*—Stress detection is an emerging topic of interest because stress is a person's affective state that have severe implications on the health. Also, wearable technology has improved, including more sensors and being less intrusive. For that reason, some researchers are focused on detecting stress using wearable devices, analyzing physiological signals to be more reliable the detection. In this paper, we show the results of our implementation of stress detectors based on two recent works that use the WESAD dataset to train and test different classifiers for detecting stress in humans.**

***Index Terms*—Stress detection, physiological signal, wearable device.**

I. Introduction

Affective computing is an emerging field to improve human-computer interaction by building empathic machines. Empathic machines detect the affective state of a human user and adapt their "behavior" accordingly. Based on this point of view, stress is an interesting affective state because of the effects of long-term stress, which can range from headaches and troubled sleeping to an increased risk of cardiovascular diseases. In this way, to build a reliable stress detector, it is important to understand that stress is primarily a physiological response to a stimulus, triggered by the sympathetic nervous system (SNS). During this response, it is increased breathing/heart rate and muscle tension; these physiological changes prepare the organism for a physical reaction.

In recent years, the specificity of the physiological responses to stress and emotional stimuli was utilized to train machine learning models to predict the affective state of a subject. Furthermore, wearable electronic devices are becoming more popular among users. These devices can be used to track steps and monitor other physical activities. Stress detection using sensors of wearable devices has recently been studied widely, and some works have been shown that stress can be recognized with quite reliably.

In this paper, we compare two works which propose stress detectors based on physiological signals recorded by wearable devices. The first work [1] proposes an open dataset named WESAD and evaluates different classifiers that has as input the extracted features from the recorded physiological signals. By the other side, the second work [2] analyzes the possibility of using sensors of smartwatches to detect stress; for that reason, the signals from wrist recorded in the WESAD dataset are combined to determine which combination of sensors (presented commonly in smartwatches) has the highest accuracy. We have implemented these works using Python 3 and the source code can be found in our GitHub repository[1].

II. WESAD Dataset

The selected papers work with the WESAD dataset, which is an open dataset proposed by Schmidt et al. [1] in 2018. This dataset contains data from 15 subjects measured using Empatica E4-wristband [3] (see Figure 1) and chest-worn RespiBAN Professional device [4] (see Figure 2). Each device includes different sensors with their corresponding frequencies; for instance, Empatica E4 includes accelerometers (ACC, 32 Hz), sensors to measure skin temperature (TEMP, 4 Hz), electrodermal activity (EDA, 4 Hz) and blood volume pulse (BVP, 64 Hz). In the same hand, RespiBAN records signals at 700 Hz, such as ACC, respiration (RESP), electrocardiogram (ECG), EDA, electromyogram (EMG), and TEMP.

Fig. 1. Empatica E4-wristband device

In the data gathering, data from three different affective states (stress, amusement, and relaxed) were collected. Approximately the length of the stressed situation was 10 minutes, amused situation 6.5 minutes and, relaxed situation was 20 minutes.

The work of Siirtola [2] studied this problem as binary classification (stressed vs. non-stressed), based only on Empatica E4 signals, where amusement and relaxed states were combined as one. In the other hand, Schmidt et al. [1] work with 2-classes (stressed vs. non-stressed) and 3-classes (stress,

[1]GitHub repository: https://github.com/amayhuaq/unsa-phd-courses/tree/master/pattern-recognition

	Feature	Description
ACC	$\mu_{ACC,i}$, $\sigma_{ACC,i}$ $i \in \{x, y, z, 3D\}$ $\|\| \int_{ACC,i} \|\|$ $i \in \{x, y, z, 3D\}$ $f^{peak}_{ACC,i}$ $i \in \{x, y, z\}$	Mean, STD for each axis separately and summed over all axes Absolute integral for each/all axes Peak frequency for each axis i
ECG and BVP	μ_{HR}, σ_{HR} Σ^f_x $x \in \{ULF, LF, HF, UHF\}$	Mean, STD of the heart rate (HR) Σ the freq. components in ULF-UHF
EDA	μ_{EDA}, σ_{EDA} $\min_{EDA}$, $\max_{EDA}$ ∂_{EDA}, $range_{EDA}$ μ_{SCL}, σ_{SCL}, σ_{SCR} $corr(SCL, t)$ $\#_{SCR}$ Σ^{Amp}_{SCR}, Σ^t_{SCR} $\int_{SCR}$	Mean, STD of the EDA signal Min and max value Slope and dynamic range Mean, STD of SCR/SCL Correlation between SCL and time # identified SCR segments Σ SCR startle magnitudes and response durations Area under the identified SCRs
EMG	μ_{EMG}, σ_{EMG} $range_{EMG}$ $\|\| \int_{EMG} \|\|$ $\bar{\pi}_{EMG}$ f^{peak}_{EMG} $\#^{peaks}_{EMG}$ μ^{Amp}_{EMG}, σ^{Amp}_{EMG} Σ^{Amp}_{EMG}	Mean, STD of EMG signal Dynamic range Absolute integral Median of the EMG signal Peak frequency # peaks Mean, STD of peak amplitudes Σ of peak amplitudes
RESP	μ_{RESP}, σ_{RESP} $\min_{RESP}$, $\max_{RESP}$	Mean, STD of RESP signal Min and max RESP value
TEMP	μ_{TEMP}, σ_{TEMP} $\min_{TEMP}$, $\max_{TEMP}$ $range_{TEMP}$ $partial_{TEMP}$	Mean, STD of the TEMP Min and max TEMP value Dynamic range Slope

TABLE I

LIST OF EXTRACTED FEATURES BASED ON [1]. ABBREVIATIONS: #=NUMBER OF, Σ=SUM OF, STD=STANDARD DEVIATION

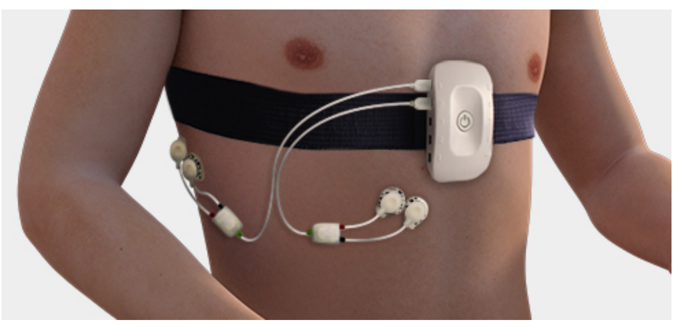

Fig. 2. RespiBAN Professional device

amusement, and relaxed) using the recorded signals from the chest and wristband. In this paper, we have focused on binary classification with the purpose to compare the results of both works.

III. Methods

Our implementation is based on Python 3 using some libraries such as sklearn and Neurokit [5] for processing physiological signals. We have implemented the functions to extract features from the signals and configured the classifiers according to the selected papers. Both works use similar features (see Table I); however, they combine signals in different ways and apply different window sizes for the training phase.

A. Work of Schmidt et al. [1]

These authors work only with a window size of 60 seconds and a slide of 0.25 seconds for the model training. This work evaluates five learning algorithms to model the classifier, such as decision tree (DT), random forest (RF), AdaBoost (AB), linear discriminant analysis (LDA) and k-nearest neighbor (kNN). This work applies different combinations of features from all signals of WESAD (chest and wrist).

B. Work of Siirtola [2]

For the model training, this author divided the signals into windows, and from these windows were extracted the features. He analyzes and compares the effect of different window sizes (15s, 30s 60s, 90s, and 120s), using a slide of 0.25 seconds. Furthermore, he evaluates the effect of combining features from different signals.

This author used 3 different classifiers: LDA, RF and quadratic discriminant analysis (QDA) for this binary classification problem.

IV. Results

We used accuracy as evaluation metric. Accuracy represents the number of correctly classified instances out of all samples. Furthermore, all models were evaluated using the leave-one-subject-out (LOSO) cross-validation (CV) procedure. This means that one person's data in turn was used for testing and other 14 for training.

According to Siirtola, the best window size is determined by the best accuracy of testing LDA classifier with all bio-signals, applying the different window sizes in the training set. Table II presents the accuracies values for each window size, and we can see that 120 seconds is the best, with the highest accuracy value (79.2%). Using this window, we analyze LDA, QDA and

RF with different sensor combinations, the results presented in Table III show our average accuracy obtained after testing these classifiers; according to these results, the best classifier is LDA, with an average accuracy of 81.2%.

Window size	Accuracy
15	76.4
30	75.3
60	76.8
90	78.4
120	**79.2**

TABLE II

WINDOW SIZE HAS AN IMPORTANT EFFECT ON THE RECOGNITION ACCURACY [2]. RESULTS USING LDA CLASSIFIER AND DATA FROM ALL EMPATICA E4 SENSORS

Sensors	LDA	QDA	RF
ACC	59.6	68.0	68.0
EDA	71.0	63.0	65.1
BVP	73.2	66.5	73.0
TEMP	78.2	63.1	63.5
EDA + BVP	72.5	68.2	75.0
EDA + TEMP	76.1	77.5	67.2
TEMP + BVP	78.8	79.4	74.6
EDA + BVP + TEMP	**81.2**	69.4	**78.1**
ACC + EDA + BVP + TEMP	80.0	**72.4**	77.2

TABLE III

AVERAGE RECOGNITION RESULTS ACCURACIES USING THE CLASSIFIERS AND SENSOR COMBINATIONS PROPOSED IN [2]

In the other hand, Schmidt et al.evaluate more classifiers, combining features from RespiBAN and Empatica E4. Table IV shows the accuracies for the different sensor combinations with the five classifiers. Analyzing these results, we can see that the best accuracy (90.1%) is achieved using the LDA classifier, which has to be trained using all signals of RespiBAN (ACC, ECG, EDA, EMG, RESP and TEMP).

Sensors	DT	RF	AB	LDA	kNN
ACC wrist	61.1	66.5	**68.7**	57.0	60.8
ACC chest	68.7	69.1	**70.2**	68.9	54.0
BVP wrist	78.4	81.1	80.9	**82.4**	78.8
EDA wrist	73.5	73.8	**76.1**	75.2	69.7
TEMP wrist	65.3	64.5	64.1	**66.3**	61.2
Wrist physio	80.6	85.0	**85.1**	83.2	78.8
ECG chest	76.9	78.8	80.1	**82.4**	75.9
EDA chest	70.0	74.2	72.1	**78.5**	66.2
EMG chest	52.6	60.0	59.2	**64.2**	54.7
RESP chest	80.1	83.4	83.4	**84.8**	73.1
TEMP chest	62.1	61.4	60.4	**65.7**	54.8
Chest physio	81.3	88.8	85.7	**90.0**	77.9
All wrist	78.9	**84.3**	79.5	82.4	60.0
All chest	78.2	88.4	88.7	**90.1**	65.2
All physio	82.1	84.5	86.5	**89.2**	80.1
All signals	80.3	83.8	84.2	**89.4**	70.8

TABLE IV

RESULTING ACCURACIES OF THE BINARY CLASSIFIERS ACCORDING TO THE CONFIGURATION PROPOSED BY [1]. IN THIS EXPERIMENT, SIGNALS FROM RESPIBAN AND EMPATICA E4 WERE USED.

V. DISCUSSION

The main focus of these papers is stress detection based on physiological data. The work of Schmidt et al. [1] proposes an open dataset to be used as a benchmark dataset for stress detection; another contribution is the application of machine learning algorithms to determine which could be an adequate classifier in this type of problem. Siirtola [2] concentrates on analyzing the impact of sensor combinations and the window size in the accuracy of classifiers. Additionally, he analyzed if the EDA signal is important for stress recognition, to extend this type of detection in smartwatches.

According to our experiments and the results of the selected papers, LDA is the most recommendable classifier to binary classification for stress detection. This technique obtains the highest accuracy in both tests and depending on the window size and sensor combinations could improve. Although using chest signals has better results, the RespiBAN device is more intrusive than the E4-wristband; for that reason, Siirtola focused his work only on signals recorded by E4.

VI. CONCLUSION

Wearable technology has been improved in the last years; as consequence, new topics of research are emerging, like stress detection using wearables. Wearable technology is based on sensors and their signals have to be processed to extract relevant features according to the objective. The selected papers are focused on processing physiological signals to detect stress in subjects, and they work using the WESAD dataset which records physiological data from RespiBAN and Empatica E4 devices. We extracted features from each window of the different signals, being some features more expensive to compute than others, in time execution and calculation. Although the configuration of classifiers and sensors combinations are based on the papers, our results are different because some methods to calculate features are not explained in the papers and those were calculated based on the processing techniques of other authors. Nonetheless, we achieve the same conclusion about that LDA is the best binary classifier for this problem.

REFERENCES

[1] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, “Introducing wesad, a multimodal dataset for wearable stress and affect detection,” in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ser. ICMI ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 400–408.

[2] P. Siirtola, “Continuous stress detection using the sensors of commercial smartwatch,” in *In 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*, ser. UbiComp/ISWC ’19 Adjunct. New York, NY, USA: Association for Computing Machinery, 2019, p. 1198–1201.

[3] Empatica E4, https://www.empatica.com/en-eu/research/e4/.

[4] RespiBAN Professional, https://www.biosignalsplux.com/index.php/respiban-professional.

[5] NeuroKit.py, https://neurokit.readthedocs.io/en/latest/index.html, 2017.