Stress Detection by Machine Learning and Wearable Sensors

PRERNA GARG, TU Kaiserslautern & German Research Center for Artificial Intelligence, Germany
JAYASANKAR SANTHOSH, TU Kaiserslautern & German Research Center for Artificial Intelligence, Germany
ANDREAS DENGEL, TU Kaiserslautern & German Research Center for Artificial Intelligence, Germany
SHOYA ISHIMARU, German Research Center for Artificial Intelligence, Germany

Mental states like stress, depression, and anxiety have become a huge problem in our modern society. The main objective of this work is to detect stress among people, using Machine Learning approaches with the final aim of improving their quality of life. We propose various Machine Learning models for the detection of stress on individuals using a publicly available multimodal dataset, WESAD. Sensor data including electrocardiogram (ECG), body temperature (TEMP), respiration (RESP), electromyogram (EMG), and electrodermal activity (EDA) are taken for three physiological conditions - neutral (baseline), stress and amusement. The F1-score and accuracy for three-class (amusement vs. baseline vs. stress) and binary (stress vs. non-stress) classifications were computed and compared using machine learning techniques like k-NN, Linear Discriminant Analysis, Random Forest, AdaBoost, and Support Vector Machine. For both binary classification and three-class classification, the Random Forest model outperformed other models with F1-scores of 83.34 and 65.73 respectively.

CCS Concepts: • **Computing methodologies → Classification and regression trees**.

Additional Key Words and Phrases: Stress detection, wearable sensor, classification

ACM Reference Format:
Prerna Garg, Jayasankar Santhosh, Andreas Dengel, and Shoya Ishimaru. 2021. Stress Detection by Machine Learning and Wearable Sensors. In *26th International Conference on Intelligent User Interfaces (IUI '21 Companion), April 14–17, 2021, College Station, TX, USA*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3397482.3450732

1 INTRODUCTION

Stress is identified as one of the top ten social determinants of health disparities. Organizations such as the World Health Organisation, the American Psychological Association, and Occupational safety and health administration are raising awareness about the negative impact of stress on health and its associated costs to society [6]. Although short-term stress responses are beneficial, prolonged exposure to stress is known to cause a number of diseases, including hypertension and coronary artery disease. Furthermore, prolonged exposure to stress can lead to mental illnesses such as depression, anxiety disorders, and burnout. Therefore it is important that people are aware of stressful situations so that they can take the necessary actions to cope with them. Considering the above, there is a need for a biofeedback system that can detect stress in a timely manner and inform the individual for the proper treatment of this situation. We have developed Machine Learning classifiers, that detects stress levels in individuals using sensors which can prevent various stress related health issues.

This paper proposes a stress prediction using different Machine Learning classifiers and determining the most effective physiological features that predict stress on one of the major stress datasets. The physiological data including ECG, EDA, EMG, RESP, and TEMP sampled at 700 Hz from a chest-worn device were used for training the models.

2 RELATED WORK

Several studies have been conducted with the aim of detecting stress automatically. The data from the smartphone's built-in accelerometer is used to detect behavior that correlates with user's stress levels [2, 4]. Another important domain where stress sensing is applied is a working environment. Hoffmann *et al.* utilized an infrared thermal camera for tracking the face temperature and have investigated the relation with the stress level [3]. Koldijk *et al.* developed automatic classifiers to examine the relationship between working conditions and mental stress-related conditions from sensor data: body postures, facial expression, computer logging and physiological data (ECG and skin conductance) [5]. Many studies have been performed on this publicly available WESAD dataset. One such study has been conducted where stress is detected automatically by various Machine Learning approaches performed on WESAD Dataset using various statistical features [1].

3 METHODOLOGY

3.1 Data and Preprocessing

We utilize a multimodal physiological dataset named WESAD for the purpose of stress detection. This dataset has been introduced and made publicly available by Schmidt *et al.* [7]. This dataset is a collection of motion data and physiological data from 15 participants. The data was collected from a chest-worn device *RespiBAN Professional* and a wrist-worn device *Empatica E4*. Participants were put into various study protocol conditions such as meditation, recovery, baseline, amusement, stress and their physiological stimuli were documented [1]. The details about sensor setup, sensor placement, and the procedure followed to develop the dataset is mentioned in detail [7]. The *RespiBAN* measured three-axis acceleration (ACC), electrocardiogram (ECG), body temperature (TEMP), respiration (RESP), electromyogram (EMG) and electrodermal activity (EDA) and the signals were sampled at 700 Hz. The raw ECG and RESP signals for a particular participant are shown in Figures 1 and 2. It can be observed from Figure 2 that for stress condition, the breathing rate was higher compared to baseline and amusement conditions.

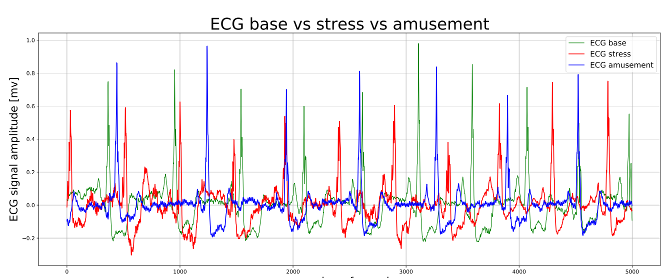


Fig. 1. Raw ECG signal for a single participant

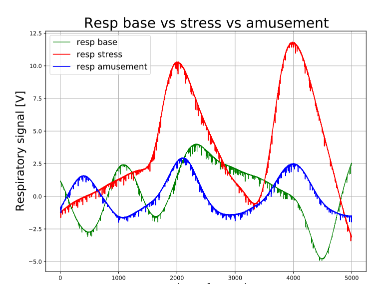


Fig. 2. Raw Respiration signal for a single participant

3.2 Feature Extraction

The raw data was analyzed and preprocessed. As mentioned by the original authors, the physiological data acquired using the chest-worn device alone provides better results than the combination of both. The dataset itself is unbalanced, with baseline or neutral condition having more samples than stress or amusement. Since the combination of features from both devices does not make much difference, the chest data has been further used in this work.

The preprocessed data was segmented using a sliding window algorithm with 10 seconds, without overlapping for all the sensor signals except ACC. The statistical features – standard deviation, mean, minimum, and maximum values – were computed for each 10-second window on raw ECG, EMG, EDA, RESP, and TEMP signals. With the sliding window approach, the raw signals were transformed into meaningful features, and they were combined to form a feature dataset. The feature dataset was further split into training and testing using *leave-one-participant-out* cross-validation, i.e., data from one participant was used for testing the model and the data from all remaining participants were used for training the model. After splitting the data into train and test, different Machine Learning algorithms were employed to train the model.

3.3 Classification

Five machine learning algorithms (Random Forest, k-Nearest Neighbor, Linear Discriminant Analysis, AdaBoost and Support Vector Machine) were used and their performances were compared. Two types of classifications were performed: three-class (neutral vs. stress vs. amusement) and binary classification (stress vs. non-stress).

The hyperparameters were tuned for the different algorithms to provide the best results. For the Random Forest classifier, minimum number of samples for splitting a node was set to five and the number of estimators was set to 50. For SVM, the radial basis function kernel was used. In the k-Nearest Neighbor algorithm, the number of neighbors was set to 100 in both classification tasks. For the AdaBoost classifier, the number of estimators was set to 25 and learning rate was set as *two*. For our work, the ground truths in the dataset used were baseline, stress, and amusement encoded as labels *zero*, *one*, and *two*, respectively. For binary classification, baseline and amusement conditions were combined into a single class (non-stress) with label *zero* and stress as another class with label *one*.

3.4 Evaluation Metric

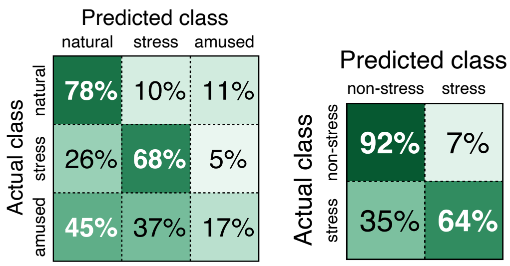
The dataset is highly imbalanced, because during the study protocol, various conditions were carried out at different lengths. Due to this imbalance in the dataset, accuracy is not chosen as the primary evaluation metric, whereas F1-score was used as the main evaluation metric. For generalization, *leave-one-participant-out* cross-validation was used for the evaluation of all models, and the final accuracy is reported as the mean of all the testing accuracies where one participant is left out for testing and others for training in each iteration. The main intention of following this procedure was to ensure that the model performed well on previously unseen participants.

4 RESULTS AND DISCUSSION

For this study, multiple classifiers were implemented for the purpose of stress detection, on the given set of participants. The performance of different classifiers based on both binary and three-class classification is shown in Table 1. For both binary classification and three-class classification, the Random Forest model outperformed other models with F1-scores of 83.34 and 65.73, respectively. The AdaBoost classifier also provided comparable results to the Random Forest classifier.

Table 1. Summary of the stress classification results using leave-one-participant-out approach

	Binary		Three class	
	F1-score	Accuracy	F1-score	Accuracy
Random Forest	**83.34**	**84.17**	**65.73**	**67.56**
Support Vector Machine	75.88	76.01	59.64	59.56
k-Nearest Neighbour	74.71	77.26	58.14	65.00
Linear Discriminant Analysis	74.70	78.47	50.44	67.06
AdaBoost	81.18	82.24	63.82	64.34



(a) Three-class classification (b) Binary classification

Fig. 3. Confusion matrices using Random Forest

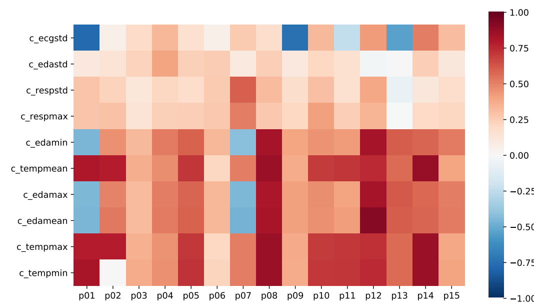


Fig. 4. Pearson Correlation between Stress and the top ten extracted features

Confusion matrices for three-class and binary classification using Random Forest were plotted. From Figure 3 (a), we observed that the three-class classification did not perform well due to the high imbalance in the data. However it is evident from Figure 3 (b), the binary classification using Random forest yield better results. The feature importance was computed using Random Forest and we observed that the standard deviation values of ECG, EDA and RESP (respiration) had the highest impact on the classification performance whereas the mean value of ECG had the least importance.

The Pearson correlation between the top ten features and stress were calculated for all the participants as shown in Figure 4. This result shows that all the participants had a positive correlation with the mean, minimum and maximum values of skin temperature. Except for participants *p01* and *p07* , all other participants had a positive correlation with the mean, minimum and maximum values of EDA. Since the dataset was highly imbalanced, it was challenging to segment the data. Furthermore, this dataset had more samples for a chest-worn device and fewer samples for a wrist-worn device, hence only samples for chest device were used in this work.

5 CONCLUSION

This study aims at implementing different Machine Learning classification models on the publicly available dataset WESAD, for the purpose of stress detection. The dataset was preprocessed, transformed, and the statistical features from the physiological signals were extracted. By using the feature dataset, multiple Machine Learning models were trained for stress detection and their performances were compared. It was observed that the Random Forest model performed better for both three-class (neutral vs. stress vs. amused) and binary classification.

As future work, we plan to extract more features from the physiological signals and implement Deep Learning models like Recurrent Neural Network for stress prediction. The self-report questionnaires filled by the participants in the dataset could be used to predict the affective state of that specific person.

REFERENCES

[1] Pramod Bobade and M Vani. 2020. Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 51–57.

[2] Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. 2015. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics* 20, 4 (2015), 1053–1060.

[3] Sabine Hoffmann, Helga Tauscher, Andreas Dengel, Shoya Ishimaru, Sheraz Ahmed, Jochen Kuhn, Carina Heisel, and Yutaka Arakawa. 2016. Sensing thermal stress at office workplaces. In *Proceedings of the 5th International Conference on Human-Environment Systems ICHES 2016 Nagoya*.

[4] Shoya Ishimaru and Koichi Kise. 2015. Quantifying the mental state on the basis of physical and social activities. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 1217–1220.

[5] Saskia Koldijk, Mark A Neerincx, and Wessel Kraaij. 2016. Detecting work stress in offices by combining unobtrusive sensors. *IEEE Transactions on Affective Computing* 9, 2 (2016), 227–239.

[6] Bhargavi Mahesh, Teena Hassan, Erwin Prassler, and Jens-Uwe Garbas. 2019. Requirements for a reference dataset for multimodal human stress detection. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 492–498.

[7] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 400–408.