Analysis of EmpaticaE4 Sensors for Stress Detection

April 19, 2023

Registration number:	2204551
Link to GitHub:	https://github.com/Socret360/ce888-data-science-and-decision-making/tree/mai

Executive summary (max. 200 words)	195
Main findings and Discussion (max. 600 words)	599
Conclusions (max. 300 words)	270
Total word count	1065

Contents

Abstract

Stress has short and long-term effects on an individual's health [7], [8]. Advancements in wearables that can monitor human motion and biosignals have led to research into their application as a non-intrusive technique for stress detection [2], [5], [6], [9], [10]. Among those devices, the EmpaticaE4 smartwatch has been the most popular device that outputs accelerometer (ACC), skin temperature (TEMP), heart rate (HR), blood volume pulse (BVP), inter-beat interval (IBI), and electrodermal activity (EDA). This report analyses the potential of these signals for stress detection by training and evaluating the F1 score of a Decision Tree (DT) classifier on the Stress-Predict dataset [6] using different sensor combinations. It demonstrates that biological signals from IBI and HR play an important role in detecting stress. More specifically, the result shows that an individual's average heart rate (μ_{hr}), heart rate standard deviation (σ_{hr}), and heart rate variability (hrv) [1] are significant factors in the model's decision. Furthermore, the report also shows that the window size for feature extraction of time series data affects the classifier's performance. In particular, increasing window size improves the classifier's performance as the model is trained on aggregated data with more information. These findings serve as the basis for developing a stress detector for wearable devices.

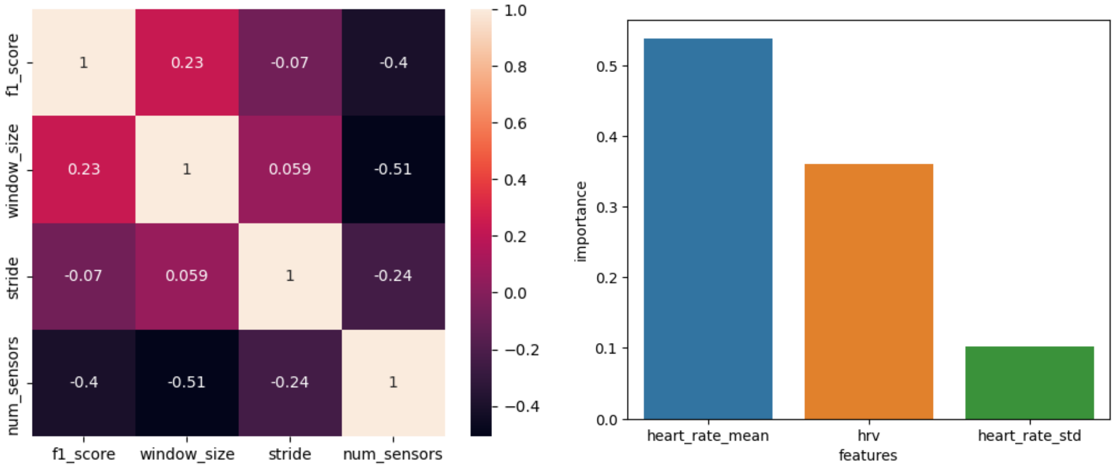
1 Main Findings

Window Size (sec)	**Stride (sec)**	**Features**	**Max Depth**	**F1 Score**
60	1	IBI+HR	5	64.23 (±0.0)
40	2	IBI+TEMP+HR	5	64.01 (±0.0)
40	2	TEMP+HR	5	62.46 (±0.0)
10	2	EDA+HR	5	62.02 (±0.0)
10	2	IBI+EDA+HR	5	61.6 (±0.0)
40	1	IBI	5	60.61 (±0.0)
20	3.5	IBI+TEMP+EDA	10	60.44 (±0.03)
20	3.5	TEMP+EDA	10	60.05 (±0.05)
20	3.5	EDA	5	59.7 (±0.0)
20	1	ACC+IBI+EDA	5	59.3 (±0.0)

Table 1: Top 10 Decision Tree and Data Preprocessing Configurations.

Sensors	**F1 Score**	**Sensors**	**F1 Score**
ACC+IBI+TEMP+EDA+HR	47.55 (±0.0)	IBI+HR	64.23 (±0.0)
IBI+TEMP+EDA+HR	54.17 (±0.93)	TEMP+HR	62.46 (±0.0)
ACC+IBI+TEMP+HR	49.66 (±1.26)	EDA+HR	62.02 (±0.0)
ACC+IBI+TEMP+EDA	49.27 (±0.0)	TEMP+EDA	60.05 (±0.05)
ACC+TEMP+EDA+HR	47.55 (±0.0)	ACC+EDA	59.18 (±0.0)
ACC+IBI+EDA+HR	43.14 (±0.0)	IBI+EDA	58.67 (±0.0)
IBI+TEMP+HR	64.01 (±0.0)	IBI+TEMP	57.78 (±0.52)
IBI+EDA+HR	61.6 (±0.0)	ACC+IBI	54.19 (±0.3)
IBI+TEMP+EDA	60.44 (±0.03)	ACC+HR	49.16 (±1.23)
ACC+IBI+EDA	59.3 (±0.0)	ACC+TEMP	46.76 (±0.84)
ACC+IBI+HR	58.19 (±0.13)	IBI	60.61 (±0.0)
TEMP+EDA+HR	54.9 (±0.09)	EDA	59.7 (±0.0)
ACC+TEMP+HR	50.26 (±0.55)	TEMP	55.46 (±0.09)
ACC+IBI+TEMP	48.52 (±1.23)	HR	53.04 (±0.13)
ACC+TEMP+EDA	45.45 (±0.02)	ACC	48.87 (±1.45)
ACC+EDA+HR	43.14 (±0.0)		

Table 2: Test Results of Models Trained Using Different Sensor Combinations



(a) (b)

Figure 1: (a) Correlations between Data Parameters and F1 Score. (b) Feature Importance of Decision Tree Trained with IBI+HR, window size 60 seconds, and stride size of 1 second.

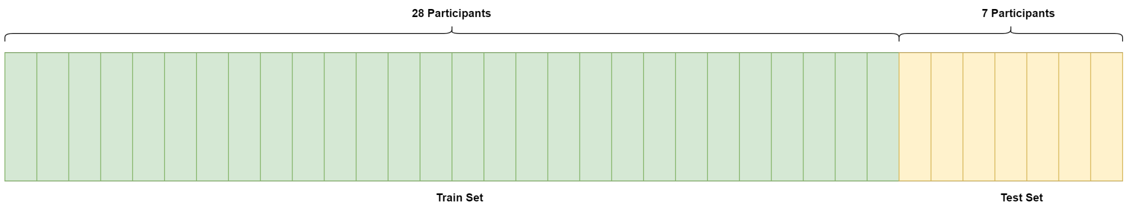


Figure 2: Train/Test Split Strategy on the Stress-Predict Dataset [6]

The experiment conducted for this paper performed leave-one-participant-out cross-validation of a DT on the training set with a total of 5425 configurations. Configurations that use the BVP signal is removed due to redundancy with the IBI and HR signals since they are derived from it [4], [3]. The best configuration of each sensor combination is then tested on the test set ten times to get its mean and standard deviation on the test set.

The result yields five findings on the use of sensors. First, according to Table 1, the best performing model uses a combination of features from the IBI and HR sensors. Among the top 10 configurations shown in Table 1, only the last one uses motion features from the ACC sensor. Second, Figure 1b shows that an individual's μ_{hr} and hrv [1] are the most significant factors impacting the model's decision, with their σ_{hr} having the least impact. Third, when used individually, Table 2 shows that IBI and EDA signal produces the highest score among other sensors, but combining them reduces the F1 score. Fourth, combining features from the HR sensor with other biological sensors (IBI, TEMP, and EDA) results in models with an increased F1 score. Finally, Figure 1a shows that increasing the number of sensors reduces the performance of the model.

There are three notable findings on window size and strides. First, the best model that combines the IBI and HR features uses a window size of 60 seconds and a stride of 1 second. Furthermore, Figure 1a shows that the window size positively correlates with the F1 score. Second, the same figure also shows that window size negatively correlates with the number of sensors used. Third, stride size does not have a clear correlation with the F1 score. However, like window size, it has a negative relationship with the number of sensors.

Overall, processing the biosignals from IBI and HR using a window size of 60 seconds and stride size of 1 second resulted in a classifier with the best performance. These findings highlight the role of biosignals in predicting stress.

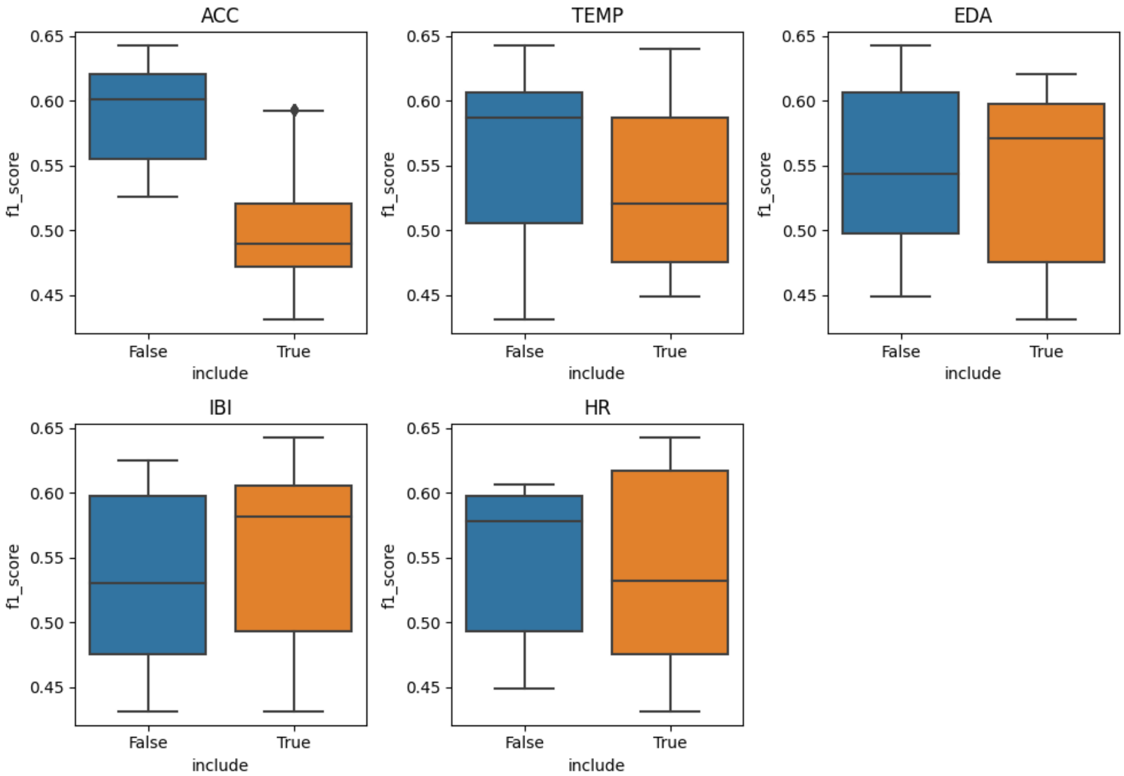


Figure 3: Inclusion of Sensors

2 Discussion

The findings from previous sections demonstrated three notable characteristics. First, it demonstrates that biosignals are more helpful in predicting stress than signals from motion sensors. This is in line with other studies that assert that stress is a response to biosignals [2], [5], [6], [9], [10]. Figure 3 shows that, on average, models using the ACC sensor perform worse than those without. One reason is that models trained with the ACC sensor learned the participant's movement during the stress and non-stress period rather than the underlying responses to stress [10]. Second, not all biosignals should be used. This is because some features from the different sensors are correlated, which is why combining them led to a reduction in performance (e.g. IBI+EDA, IBI+TEMP). It also explains the negative correlation between the number of sensors with the F1 score in Figure 1a. Third, using a larger window size improves the model's performance, but only when the number of sensors used is small (≤ 2). It may be due to the effects of redundant features from different sensors mentioned previously. It may also suggest that more fine-grain aggregation is needed when features from more sensors are added.

While the result demonstrates the potential of biosignals as beneficial for predicting stress, it has a few limitations. First, it does not show if this combination of sensors (IBI + HR) provides the best predictions on an individual level. Second, the experiment utilises a fixed window size and stride for feature extraction across all the sensors, which may only be suitable for some sensors. Third, the number of features created for each sensor is small compared to other studies [2], [5], [6], [9], [10].

3 Conclusions

This paper reinforces the findings of other studies, which show that stress is a response to biosignals. It has demonstrated that the features from IBI (hrv) and HR (μ_{hr}, σ_{hr}) sensors play a crucial role in determining stress. Furthermore, while the resulting F1 score of the best DT configuration is not high, performing the experiment using only DT allows the experiment to include more configurations (5425 configurations) during the hyper-parameter tuning and cross-validation process. A large number of configurations would not be possible with more complicated models (e.g. Random Forest and SVM). In addition, it also emphasizes the importance of feature extraction when dealing with signals from the EmpaticaE4. In particular, special care must be put into determining the best window size since bigger window sizes do not automatically result in better performance.

It is also beneficial to explore three other directions for further improvements. First, individual-based feature selection can be introduced to produce a personalised stress detection model. Siirtola and Roning [10] shows that this approach is more suitable than training a personalised model for an individual using just his/her own data. Second, an experiment with different window sizes for different sensors can be explored. Schmidt et al. [9] adopted a window size of 5 seconds for the ACC sensor and 60 seconds for the biosignals (IBI, BVP, TEMP, and EDA). Finally, more complicated features can be extracted from the time series data. In particular, Iqbal et al. [6] have demonstrated that respiratory rate can be a useful stress indicator.

Overall, this paper supports the exploration of biosignals (HR and IBI) to build features for a stress detector to be deployed on wearable devices.

References

[1] Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. 93(5):1043–1065.

[2] V. Chandra, A. Priyarup, and D. Sethia. Comparative Study of Physiological Signals from Empatica E4 Wristband for Stress Classification. In M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, T. Ören, and V. R. Sonawane, editors, *Advances in Computing and Data Sciences*, volume 1441, pages 218–229. Springer International Publishing.

[3] Empatica Support. E4 data - HR.csv explanation.

[4] Empatica Support. E4 data - IBI expected signal.

[5] T. Iqbal, A. Elahi, S. Ganly, W. Wijns, and A. Shahzad. Photoplethysmography-Based Respiratory Rate Estimation Algorithm for Health Monitoring Applications. 42(2):242–252.

[6] T. Iqbal, A. J. Simpkin, D. Roshan, N. Glynn, J. Killilea, J. Walsh, G. Molloy, S. Ganly, H. Ryman, E. Coen, A. Elahi, W. Wijns, and A. Shahzad. Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset. 22(21):8135.

[7] S. J. Lupien, B. S. McEwen, M. R. Gunnar, and C. Heim. Effects of stress throughout the lifespan on the brain, behaviour and cognition. 10(6):434–445.

[8] A. S. Nagaraja, N. C. Sadaoui, P. L. Dorniak, S. K. Lutgendorf, and A. K. Sood. SnapShot: Stress and Disease. 23(2):388–388.e1.

[9] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. pages 400–408. ACM.

[10] P. Siirtola and J. Röning. Comparison of Regression and Classification Models for User-Independent and Personal Stress Detection. 20(16):4402.

Appendix

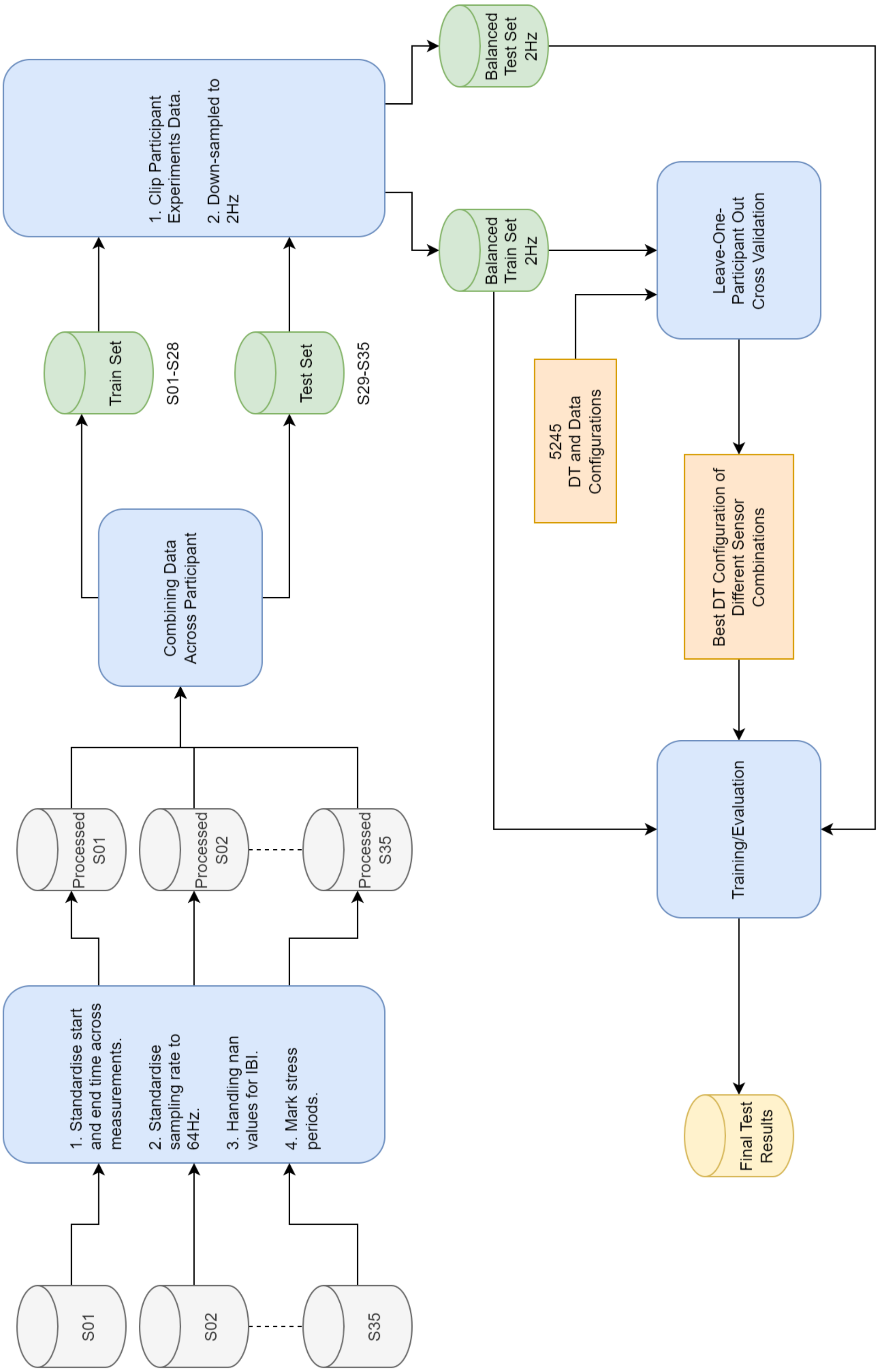


Figure 4: Experiment Process