# Aware of Your Stress: A Stress Detection Using Wearable Watch

April 23, 2023


| | |
|---|---|
| Registration number: | 2200324 |
| Link to GitHub: | https://github.com/SuwitchayaK/Stress_Detection.git |


| | |
|---|---|
| Executive summary (max. 200 words) | 163 |
| Main findings and Discussion (max. 600 words) | 594 |
| Conclusions (max. 300 words) | 220 |
| Total word count | 977 |

## Contents

**Abstract**

The wearable watch is an accessible and user-friendly device. It may have the potential to accurately predict stress, especially for patients who are more sensitive to stress. Ensuring prediction accuracy is crucial for the device as it can prevent harm to patients and help them manage their stress levels. In order to detect stress, four machine learning algorithms including Random Forest, Support Vector Machine, K-Nearest Neighbors, and Ridge Logistic Regression are used. I found that the watch can detect stress with available sensors. However, the result is still not adequate. The Random Forest algorithm has the highest predictive power compared to other algorithms. I also found that blood volume pressure has the highest contribution to predicting stress, and including 15-second lag variables improves accuracy. With the unbalanced dataset, the stress period may be underrepresented and cause a False Negative problem. I found that allowing the Random Forest algorithm to predict more positive result is the most effective method to cope with this problem.

# 1 Main Findings

## 1.1 Data

The data is retrieved from open source [5, 4] for 25 participants. There are 6 features collected from the watch: heart rate (HR), blood volume pressure (BVP), accelerometer (ACC), skin conductance (EDA), inter-beat interval (IBI), and skin temperature (TEMP). ACC is the overall magnitude and IBI is generated using sliding windows of 5 seconds. Each feature is processed to 1 Hz frequency. The missing values are filled using means between the time difference. All features are rescaled as they have different units. The watch recorded through non-stress and stress periods. These periods are converted to binary values. The distribution of each period is shown in Figure 1.

The data was split for 80% for the train set and 20% for the test set. All participants are added into both sets in temporal orders.

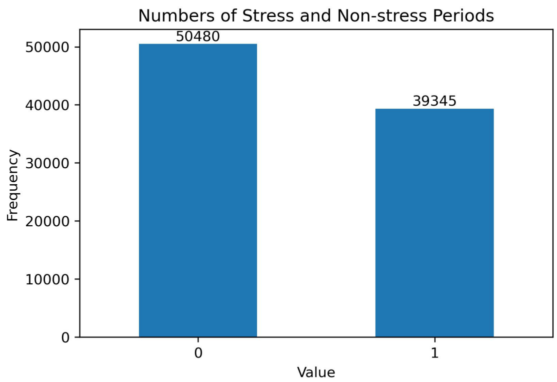


Figure 1: Stress and Non-stress Periods

## 1.2 Analysis

| Features | Features Scores |
|---|---|
| BVP | 0.45 |
| TEMP | 0.15 |
| EDA | 0.15 |
| ACC | 0.11 |
| HR | 0.10 |
| IBI | 0.04 |

Table 1: Features Contributions

The Random Forest (RF) Algorithm is used for feature extraction. Table 1 shows that BVP has the highest contribution to stress, while the others do not show obvious contributions.

The data may have temporal dependencies. Accordingly, I created lag variables of 10, 15, and 30 seconds for each feature to train the models. By using Random Forest (RF) model, adding 15-second lag variables gives the strongest predictive power as shown in Table 2.

| Lags | Accuracy Rate |
|---|---|
| 10 | 0.45 |
| 15 | 0.49 |
| 30 | 0.43 |

Table 2: Accuracy rates for each time lag

| Models | Accuracy Rate | F1(non-stress) | F1(stress) | FNR |
|---|---|---|---|---|
| RF | 0.49 | 0.49 | 0.43 | 0.59 |
| RF balanced class weight | 0.50 | 0.61 | 0.30 | 0.76 |
| RF synthetic sample | 0.49 | 0.61 | 0.28 | 0.77 |
| RF reduce decision threshold | 0.46 | 0.39 | 0.53 | 0.33 |
| RF Selected Features | 0.45 | 0.43 | 0.48 | 0.70 |
| SVM penalty=1 | 0.32 | 0.39 | 0.23 | 0.77 |
| SVM penalty=2 | 0.32 | 0.32 | 0.33 | 0.63 |
| l2 Logit | 0.57 | 0.72 | 0.10 | 0.95 |
| l2 Logit reduce decision threshold | 0.58 | 0.71 | 0.30 | 0.80 |
| KNN | 0.45 | 0.54 | 0.31 | 0.72 |

Table 3: Results from Models

There are algorithms that are commonly used for stress detection including RF, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)[3]. I used those algorithms and Ridge (l2) Logistic Regression (Logit) as it helps regularize the data and extracts important features. The lag variables are included in each model.

The numbers of stress and non-stress periods are unbalanced which can cause misclassification for the underrepresented group. There are three methods to mitigate the problem for the RF algorithm: balanced class weight to make the model more sensitive to the stress period, create synthetic sample for the stress period and reduce decision threshold to make the model more sensitive to predict positive results. For the SVM model, I increased penalty for misclassification, however, this may cause overfitting problem. For KNN, I chose 2-Nearest Neighbors for prediction. Lastly, the selected features model includes BVP, TEMP, and EDA using RF. In addition, I included False Negative Rate (FNR) from each model.

Table 3 shows that the Ridge Logistic Regression with a reduced decision threshold has the highest accuracy to classify stress. However, it cannot classify well when the participant is stressed. Comparatively, the RF model with a reduced decision threshold has a lower accuracy rate but has the lowest FNR. To minimize the False Negative problem, the Random Forest with a reduced decision threshold may be the best model.

### 1.3 Validation

| Model | Average Accuracy Rate |
|---|---|
| RF | 0.67 |
| l2 Logit | 0.62 |
| SVM | 0.59 |
| KNN | 0.53 |

Table 4: 10-Fold Cross Validation Scores

In order to perform 10-Fold Cross Validation, I used time-series splits to maintain the temporal orders of the sample. Table 4 shows that the RF is the best model to predict stress. The current device can predict stress. However, the results are inadequate.

## 2 Discussion

For the main result, all variables are included since it slightly improves accuracy scores and reduces False Negative. The lag variables cannot be generated in further time as more sacrifice of data. During Cross Validation process, the data is very small when it is split by participant and time so it causes low accuracy rate. Hence, I used the cross-validation only by time. There are many missing values of IBI and they are filled using interpolation and backward fill. That may cause lower contribution to the model. The endings of the last stress periods are estimated as the data is unavailable. The environment is rather relaxed during the stress experiment [5]. Hence, the difference between stress and non-stress periods may not be distinguished.

# 3 Conclusions

To conclude, the current wearable watch can detect stress. However, the result may not be adequate. By using the Random Forest algorithm, I found that blood volume pressure has the highest contribution to stress, and including 15-second lag variables improved detection accuracy. Four machine learning algorithms are used to classify stress including Random Forest, Support Vector Machine, K-Nearest Neighbors, and Ridge Logistic Regression. The Random Forest algorithm generates more valid results compared to other algorithms. In addition, the unbalanced number of stress and non-stress periods may induce inaccuracy when predicting stress. I used different methods to mitigate the problem. I found that the Random Forest algorithm effectively reduces False Negative when reducing the prediction threshold or making it more sensitive to predict the positive results.

For further improvements, the company may consider having experiments for longer periods, improving the algorithm individually, and detecting specific movements instead of overall magnitude. The previous study [6] showed that the RF is more predictive for 15 minutes before physical symptoms of stress occur, and each individual has specific features and algorithms to detect stress. It also suggested that detecting specific movements that can be signs of stress may improve prediction. I also encourage the company to expand the experiment on different age groups and genders as they variously perceived stress [1, 2] to improve predictability.

# References

[1] C. A. Eisenbarth. Coping with stress: Gender differences among college students. *College Student Journal*, 53(2):151–162, 2019.

[2] S. Folkman, R. S. Lazarus, S. Pimley, and J. Novacek. Age differences in stress and coping processes. *Psychology and aging*, 2(2):171, 1987.

[3] S. Gedam and S. Paul. Automatic stress detection using wearable sensors and machine learning: A review. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–7, 2020.

[4] T. Iqbal, A. Elahi, S. Ganly, W. Wijns, and A. Shahzad. Photoplethysmography-based respiratory rate estimation algorithm for health monitoring applications. *Journal of medical and biological engineering*, 42(2):242–252, 2022.

[5] 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. *Sensors*, 22(21):8135, 2022.

[6] C. W. Jin, A. Osotsi, and Z. Oravecz. Predicting stress in teens from wearable device data using machine learning methods. 2020.