Detecting Stress with Wearable Watch Sensors: A Data Science Investigation of Nurse Stress Data Set

April 23, 2023

Registration number: 2209337
Link to GitHub: https://github.com/TNONTANT/Detecting-Stress-with-Wearable-Watch-Sensors

Executive summary (max. 200 words)	184
Main findings and Discussion (max. 600 words)	597
Conclusions (max. 300 words)	221
Total word count	1002

Contents

Abstract

Stress is a pervasive issue that affects millions of people worldwide, with negative consequences on both physical and mental health. Wearable devices, such as smartwatches, have the potential to monitor and detect stress in real-time, enabling individuals to take timely action and improve their well-being. In this study, we explore the potential of developing a stress detection feature for wearable watches, using the Nurse stress dataset.

By analyzing data collected from sensors, identifying patterns, and building a machine learning model, researchers were able to achieve impressive results. To boost accuracy, they applied the SMOTE upsampling technique, resulting in even higher recall. The study revealed that the random forest classifier and XGBoost classifier were the top-performing models, with accuracy and recall rates of up to 92.57% and 84.78%, respectively. The analysis also identified skin temperature, electrodermal activity, and inter-beat interval as the key factors affecting stress detection. While the study is limited by a small dataset, the findings provide a promising starting point for future research to explore larger datasets and enhance the model's generalizability, ultimately making it a useful tool for real-world stress detection.

1 Main Findings

1.1 Methodology

In this study, we utilized the A multimodal sensor dataset for continuous stress detection of nurses in a hospital dataset [1] to develop a real-time stress detection algorithm. The dataset includes biometric data and stress survey responses from 15 nurses during their working hours.

1.1.1 Signal Analysis

The examination of figures 1 and 2 reveals that signals associated with varying levels of stress display unique patterns of signal.

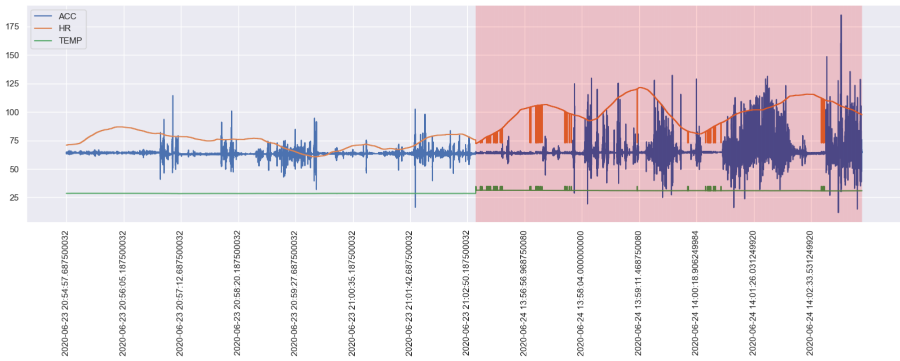


Figure 1: Plot of acceleration(ACC), heart rate(HR) and skin temperature(TEMP) signal with time, red highlight show when stress level is 2 and no colour when stress level is 0.

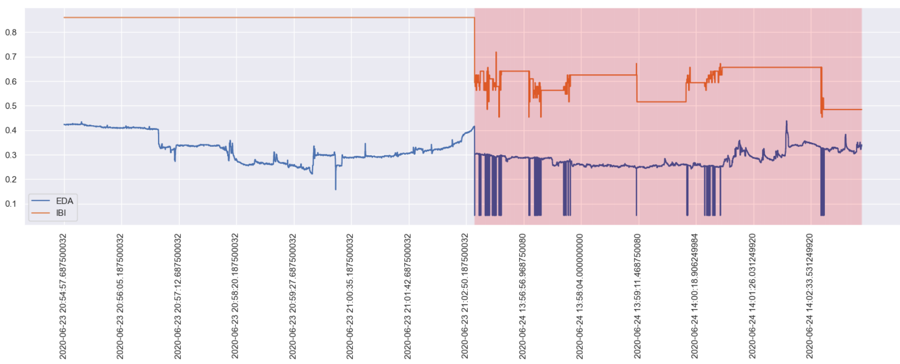
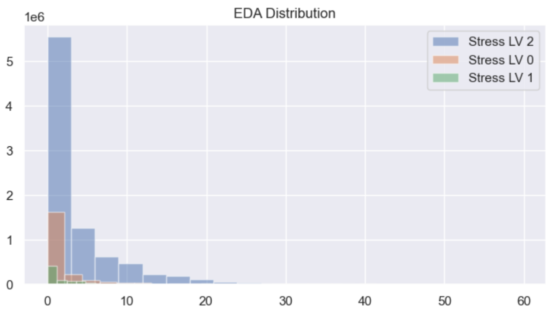
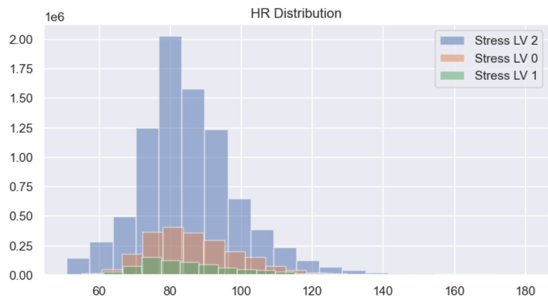


Figure 2: Plot of inter-beat interval(IBI) and electrodermal activity(EDA), red highlight show when stress level is 2 and no colour when stress level is 0.

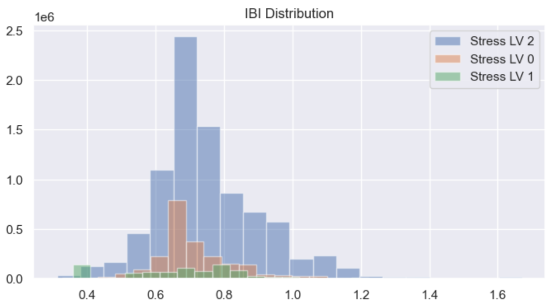
Figure 3 displays the distinctive distribution forms that differentiate the signal patterns and trends across different stress level classes. Meanwhile, Table 1 presents the corresponding statistical values for these distributions. Based on the table, EDA, IBI and TEMP signals exhibit an increasing trend as the stress level rises, as evidenced by their higher mean values, while HR shows a decreasing trend.



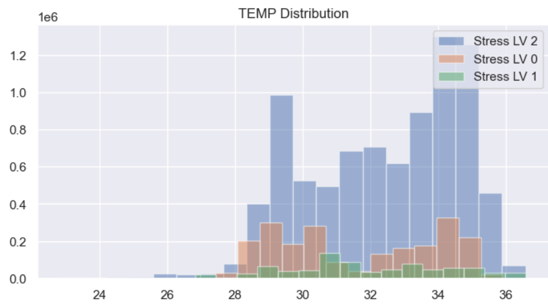
(a) Electrodermal activity distribution plot



(b) Heart rate distribution plot



(c) inter-beat interval distribution plot



(d) Skin temperature distribution plot

Figure 3: distribution of each signal of interest and examine how they differ across different stress levels.

Signal	Stress LV	Mean	Std	Max	Min	Median	Skewness	Kurtosis
EDA	stress 0	2.9472	6.4708	43.5634	0.0000	0.4641	3.6099	13.6858
	stress 1	3.0656	4.6852	24.1673	0.0000	1.0870	2.3191	5.1540
	stress 2	3.6768	5.4976	59.7607	0.0000	1.5487	2.8300	11.5828
TEMP	stress 0	31.6270	2.3430	35.9100	22.8500	31.0900	-0.0162	-1.3275
	stress 1	31.9899	2.3334	36.5900	24.3900	31.5900	0.0313	-0.7414
	stress 2	32.3998	2.2275	36.5700	22.8500	32.8100	-0.4384	-0.7349
HR	stress 0	86.6476	13.0195	169.9300	55.1300	84.8800	0.7786	1.1706
	stress 1	86.4017	15.4198	163.5000	55.9200	83.3800	1.0264	1.5200
	stress 2	85.4665	14.3729	180.2300	51.0000	83.6300	0.9504	2.5968
IBI	stress 0	0.7097	0.1061	1.3594	0.3281	0.6719	0.7248	1.8119
	stress 1	0.6730	0.1840	1.3750	0.3594	0.7031	-0.1737	-0.2233
	stress 2	0.7519	0.1557	1.6719	0.3125	0.7188	1.0267	2.9900

Table 1: Statistic value for EDA, TEMP, HR and IBI differ across different stress levels. levels

1.1.2 Feature Engineering

To extract meaningful insights from the raw stress signal data and minimize the impact of noise, we computed statistical values for fixed-size windows. We chose a window size of 1 minute, with non-overlapping windows to prevent any data leakage between the training and testing sets. This window size is half of the shortest period of time that participants reported as stressful in the stress survey (2 minutes). For each window, we calculated 7 statistical values[2] including mean, standard deviation, skewness, maximum, minimum, median, and kurtosis[1], resulting in a total of 56 features

[1]Kurtosis is a statistical measure of a distribution's shape. A high kurtosis value indicates that a distribution has heavy tails and a sharp peak, while a low kurtosis value indicates a flatter distribution with lighter tails.

from the 8 signals. These features capture the distribution and statistical characteristics of the raw signal data for each window, allowing for further analysis to identify trends and patterns in stress levels over time.

1.1.3 Class Balancing

Survey result shows that stress level 2 was recorded for 71% of the time, while stress levels 0 and 1 were only recorded for 17.65 and 8.33%, respectively. As a result, there is a clear class imbalance present in the data, which can hinder the performance of classification models on minority classes. To address this issue, the SMOTE[2] technique[3][4] was deployed to balance the classes in the training set.

1.1.4 Machine Learning Model

The data was partitioning into training 80% and testing set 20% then, the standard scalar was applied to normalize the data prior to training the four selected machine learning models: Random Forest Classifier, Support Vector Machine Classifier, Gradient Boosting Classifier, and Multi-Layer Perception Classifier.

1.1.5 Evaluation and optimization

We used k-fold cross-validation and evaluation metrics including accuracy, precision, recall, and f1-score. We also fine-tuned the hyperparameters of each model using a grid search approach to achieve the best accuracy scores possible. This methodology ensures the model can generalize well and remain robust under various circumstances.

1.2 Result

Model	Dataset	Accuracy	Precision	Recall	F1-Score
RandomForest Classifier (RF)	Original	86.34	95.81	61.62	70.33
	SMOTE	84.82	94.09	74.41	82.65
Support Vector Machine (SVMs)	Original	82.91	89.71	53.66	60.16
	SMOTE	84.62	74.70	70.97	72.61
XGBoost Classifier (XGB)	Original	92.57	94.17	81.56	86.72
	SMOTE	92.57	89.19	84.78	86.73
Mulilayer Perception (MLP)	Original	81.35	77.81	67.52	71.69
	SMOTE	80.97	73.61	70.90	72.06

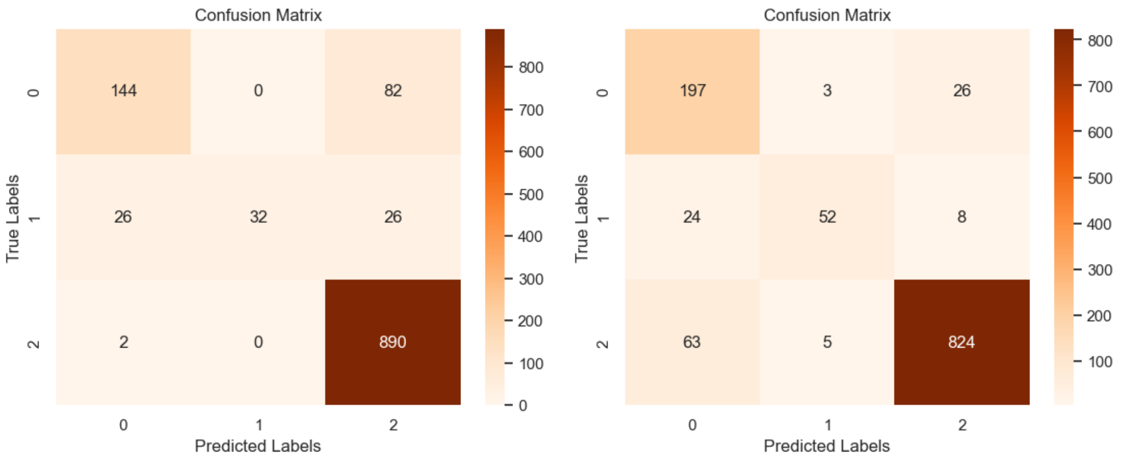
Table 2: Performance of 4 classification models compared using original and SMOTE training sets. Results based on average Accuracy, precision, recall, and F1-score from 5-fold cross-validation.

Table 2 presents the experimental results and shows that the XGBoost Classifier, when trained with SMOTE training set, achieves the best performance, with an 92.57% accuracy, 89.19% precision, recall 84.78% recall, and 86.73% f1-score. Although SMOTE improves the recall and enhances the ability of the model to predict the minority classes, there is a trade-off in terms of reduced precision, and the model performs worse on the majority class than the model trained with the original training set. This trade-off is evident from the results presented in table 2 and confusion matrix in figures 4.

2 Discussion

After analyzing the data, we found that the signals TEMP, IBI and EDA were the most important features for predicting stress levels, as indicated by their high feature importance rankings. Shows in the top 15 features based on statistical values from these three signals (as shown in Figure 5). Our model's performance was able to meet the high recall after up-sampling minority class. However, there are limitations to our methodology, particularly in regards to the generalizability of our results to other populations and contexts. We suggest further research to address these limitations and continue to improve the model's performance.

[2]1 SMOTE is a widely used algorithm for oversampling imbalanced data sets. It creates synthetic examples by interpolating between minority class samples.



(a) Model train with original training set

(b) Model train with SMOTE training set

Figure 4: Comparison of confusion matrix between 2 training set

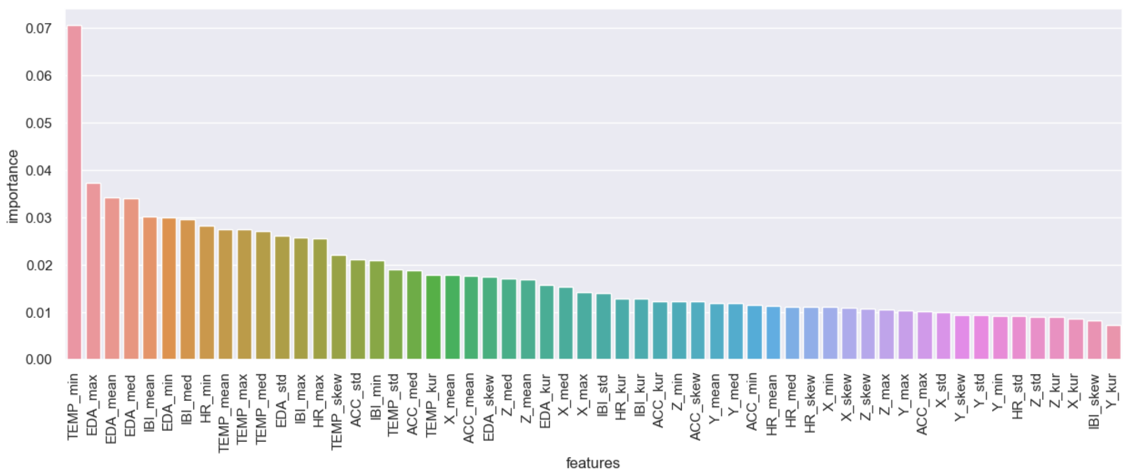


Figure 5: 56 Statistical features extracted from raw signal and ranked by importance using an Randomforest Classifier model.

3 Conclusions

In conclusion, our analysis successfully implemented machine learning techniques to predict stress levels using biometric signals. We were able to achieve a high recall rate, meeting the company's stringent requirement of low false negatives. Furthermore, the selected window size of 1 minutes allows for real-world application, as stress levels can be predicted within a practical timeframe.

However, there are limitations to our methodology that should be considered. Firstly, the data collected may not be representative of the general population, as it was collected from a specific group of people. This may limit the generalizability of our results to other populations and contexts. Secondly, while we used SMOTE to address the class imbalance issue, there is still a trade-off between precision and recall when using this technique.

To improve our methodology, future work should focus on collecting more diverse data that represents a wider range of the population. This can help to improve the generalizability of our model and increase its effectiveness in predicting stress levels in various contexts. Additionally, exploring other techniques to address the class imbalance issue, such as using different resampling methods or adjusting classification thresholds, may also improve our model's performance.

Overall, our study demonstrates the potential of machine learning techniques in predicting stress levels using biometric signals, and provides a foundation for future research in this area.

References

[1] S. Hosseini, R. Gottumukkala, S. Katragadda, R. T. Bhupatiraju, Z. Ashkar, C. W. Borst, and K. Cochran, "A multimodal sensor dataset for continuous stress detection of nurses in a hospital," *Scientific Data*, vol. 9, p. 255, 6 2022.

[2] C. Guo, M. Lu, and J. Chen, "An evaluation of time series summary statistics as features for clinical prediction tasks," *BMC Medical Informatics and Decision Making*, vol. 20, p. 48, 12 2020.

[3] M. Arslan, M. Guzel, M. Demirci, and S. Ozdemir, "SMOTE and Gaussian Noise Based Sensor Data Augmentation," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, pp. 1–5, IEEE, 9 2019.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 6 2002.