Stress Detection in Nurses by using data during the Covid Outbreak

August 23, 2023

Registration number: 2202566
Link to GitHub: https://github.com/DeeKay3745/CE888-7-SP-Data-Science-and-Decision-Making

Executive summary (max. 200 words)	124
Main findings and Discussion (max. 600 words)	587
Conclusions (max. 300 words)	171
Total word count	882

Contents

Abstract

During the challenging COVID-19 pandemic, the resilience and well-being of our front-line healthcare workers, particularly nurses, emerged as a paramount concern. Drawing from a dataset captured during this time, my research explored the stress indicators among nurses, employing an array of machine-learning techniques. Notably, the Random Forest classifier demonstrated the highest accuracy in stress detection. The findings of this study have crucial implications for organizational leadership: By integrating such predictive tools into our healthcare infrastructure, we can proactively address stress-related issues, ensuring optimal patient care and fostering a supportive work environment for our invaluable staff. These findings can be a foundation for future policies and initiatives to prioritize our healthcare professionals' mental and emotional well-being, enhancing overall organizational performance and patient satisfaction.

1 Main Findings

The dataset provides invaluable insights into pinpointing stress in complex job environments, supporting further research geared towards bolstering nurses' emotional health. Signals like electrodermal activity (EDA), Heart Rate (HR), skin temperature (TEMP), accelerometer coordinates (X, Y,Z), and blood volume pulse (BVP) were captured using the Empatica E4 wristband, which transmitted this data in real-time to a designated app on the nurse's phone.

Nurses had the option to tag undetected stress moments through the wristband, adding to the dataset. Early detection of stress is pivotal in effectively mitigating the repercussions of prolonged stress, as documented in [1].

Notably, EDA exhibited a strong correlation with most other recorded signals. The broader consensus among nurses supported the stress detection algorithm's reliability in gauging their stress levels. To sum it up, this study emphasizes the importance of real-world stress assessment and the creation of adept algorithms to facilitate early detection, with the ultimate goal of improving emotional health outcomes.

1.1 Before Pre-processing part

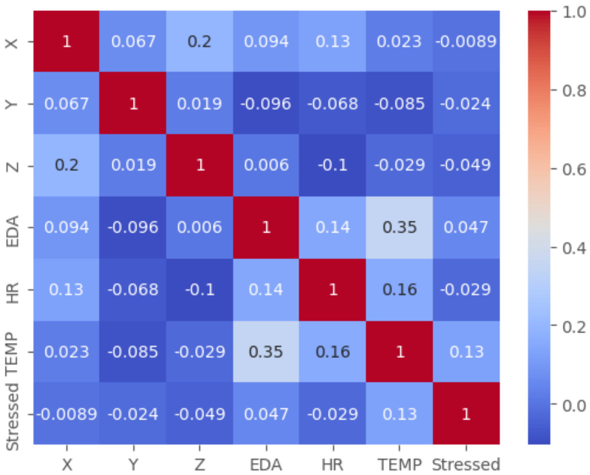
Link to original Dataset: Here[2]

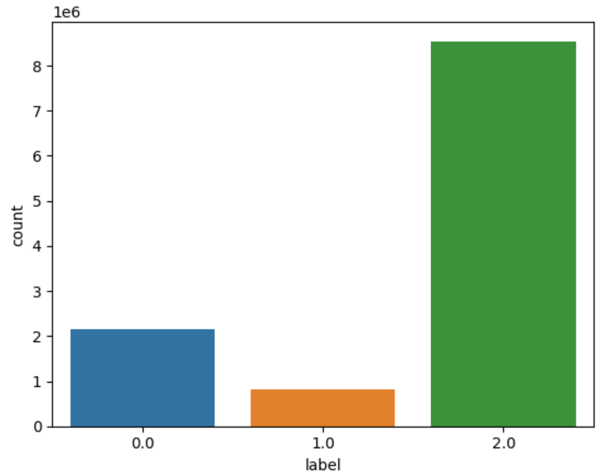
Utilizing the provided link, I accessed a dataset comprising continuous stress data from 15 nurses, collected from the beginning to the end of their shifts. The comprehensive system was engineered for real-time data collection, processing, and stress detection. The introduced data collection method, stress detection algorithm, and feedback collection tool were all tailored for instantaneous stress detection. After using the four files Prof. Ana Matran Fernandez shared, I ultimately obtained "merged_data_labeled.csv."

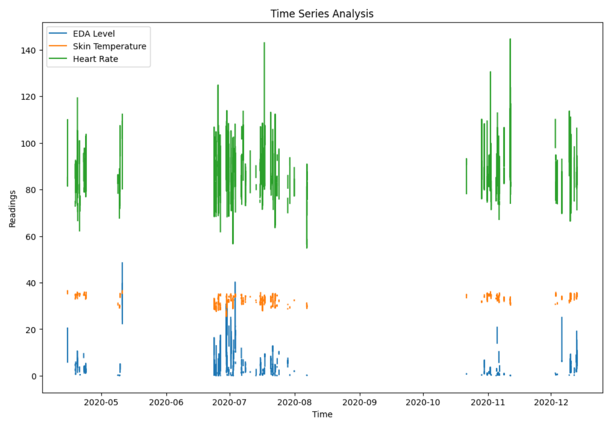
Link to final Dataset: Google Drive link

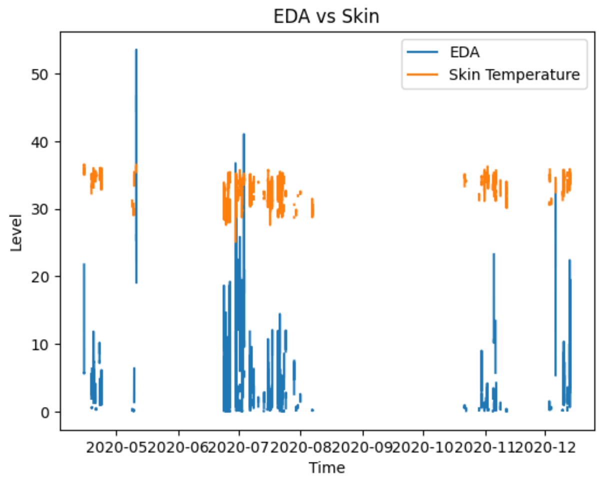
1.2 Data Exploration

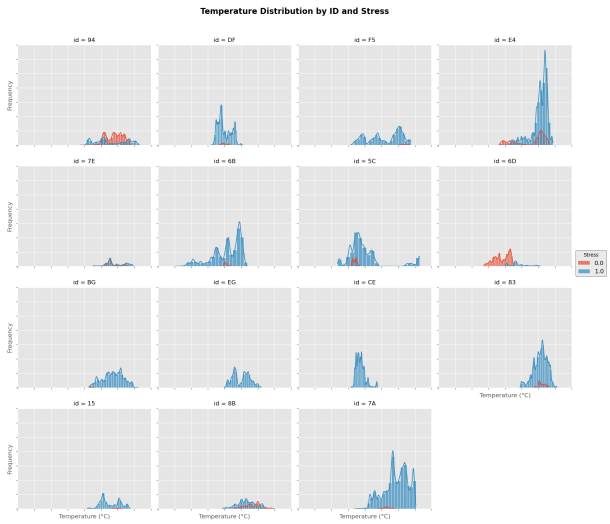
Here, are some plots that were used to understand the data better.











1.3 Data Modeling

We tested several machine learning algorithms to determine the best fit for our data:

- Random Forest Classifier
- Desicion Tree Classifier
- Logistic Regression
- Naive Bayes
- AdaBoost Classifier

Performance metrics such as accuracy, precision, recall, F1-score, R squared, and Root Mean Square Error (RMSE) were used to evaluate and compare the models. Among the algorithms mentioned above, both the Random Forest and Decision Tree classifiers achieved the highest accuracy on validation and the test set.

Label	Number of values
0.0	2162246
1.0	9346805
Total	11509051

Preprocessing :

- There were no missing values detected.
- Standardized the data using the Standard Scaler.

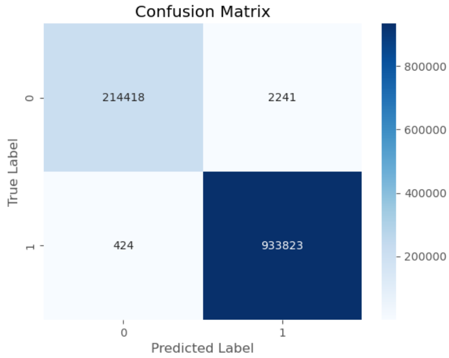


Figure 1: Confusion matrix for Random Forest Classifier

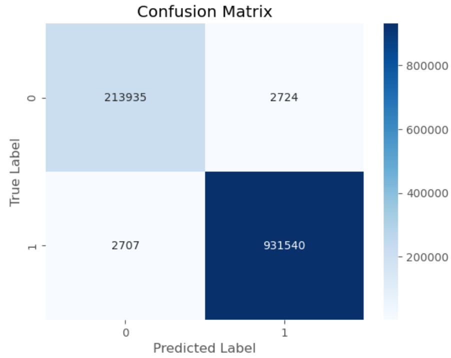


Figure 2: Confusion matrix for Decision Tree Classifier

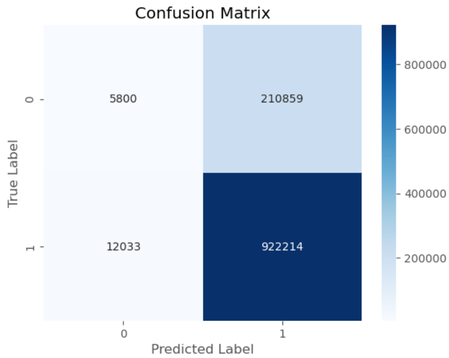


Figure 3: Confusion matrix for Naive Bayes

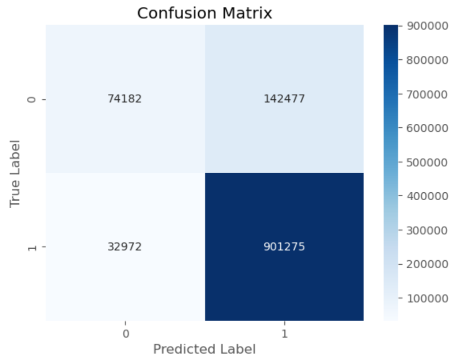


Figure 4: Confusion matrix for AdaBoost

Features :
Selected based on domain knowledge and correlation:

- Heart Rate, EDA, Temperature and Accelerometer(X, Y and Z).

For Validation set:

Algorithm	Accuracy	F1 score	Recall	Precision	R squared	RMSE
Random Forest Classification	0.9976	0.9976	0.9976	0.9976	0.9844	0.04877
Decision Tree Classifier	0.9953	0.9953	0.9953	0.9953	0.9692	0.0684
LogisticRegression	0.8121	0.7279	0.8121	0.6595	-0.2313	0.4334
Naive Bayes	0.8067	0.7342	0.8067	0.7227	-0.2669	0.4396
Ada Boost Classifier	0.8067	0.7342	0.8067	0.7227	-0.2669	0.4396

For Test set:

Algorithm	Accuracy	F1 score	Recall	Precision	R squared	RMSE
Random Forest Classification	0.9976	0.9976	0.9976	0.9976	0.9848	0.04812
Decision Tree Classifier	0.9952	0.9952	0.9952	0.9952	0.9691	0.06869
LogisticRegression	0.8117	0.7274	0.8117	0.6589	-0.2319	0.4338
Naive Bayes	0.8063	0.7335	0.8117	0.7219	-0.2673	0.4400
Ada Boost Classifier	0.8475	0.8259	0.8475	0.8312	0.0024	0.3904

2 Discussion

The Empatica E4 wristband was utilized to capture physiological data from the subject's wrist, which was promptly relayed to their smartphone via Bluetooth and then to an analytics server through Wi-Fi. Throughout their shifts, nurses wore this device for seamless real-time data gathering and stress detection[1]. An Excel file houses the participants' survey outcomes and annotated stress levels, organized by individual ID sheets. Each sheet also features a consistent ID column for ease of reference[2].

Column	Name	Description
A	ID	Anonymized identifier for the user.
B	Start time	Commencement time of the event.
C	End time	Conclusion time of the event.
D	Duration	Total time span of the event.
E	Date	The specific day when data was collected.

When examining the validation set outcomes, the Random Forest Classifier notably outperformed other algorithms, achieving near-perfect scores across all metrics, specifically around the 99.76% mark. The Decision Tree Classifier, not far behind, showcases similar efficacy with figures centered at approximately 99.53%. In contrast, Logistic Regression and Naive Bayes display a marked reduction in their performance metrics, both averaging around 81% for Accuracy and Recall, with Logistic Regression depicting a noticeably lower Precision. Ada Boost Classifier's results align closely with those of Naive Bayes.

Upon evaluating the test set results for diverse algorithms, it is evident that the Random Forest Classifier excels in accuracy and across F1 score, Recall, Precision, and R-squared metrics, all hovering around 99%. The Decision Tree Classifier mirrors this commendable performance, trailing closely. On the other hand, while Logistic Regression and Naive Bayes consistently hover around an 80% mark for Accuracy, Recall, and F1 score, their Precision tends to lag, especially in the case of Logistic Regression. Ada Boost Classifier, although lagging slightly behind in most parameters compared to Random Forest and Decision Tree, showcases a notable uplift in its performance, especially in Precision, signaling its nuanced handling of false positives.

3 Conclusions

In conclusion, the Empatica E4 wristband has showcased its capability as a premier real-time physiological data capture tool, efficiently recording parameters like EDA, Heart Rate, skin temperature, and accelerometer readings. This data, channeled instantly to the smartphone and then to an analytics server, gives us a profound insight into the nurses' stress levels during their shifts. The meticulous organization of this data in Excel, segmented by participant IDs, underscores the essence of structured data gathering in health research. The near-perfect accuracy of algorithms like the Random Forest and Decision Tree Classifiers validate the efficacy of our approach.

Given the results, it is highly recommended for CEO and decision-makers to invest further in wearable technology and enhance data analytics capabilities. This could facilitate early detection and intervention strategies for managing stress in high-pressure job roles, potentially increasing productivity and ensuring better mental health for employees. Moreover, expanding such monitoring to other organizational roles might offer a holistic view of workplace wellness and pave the way for a more holistic, employee-centric organizational culture.

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, 9(1):1–13, 2022.

[2] Ravi, M. S. Stress-detection-in-nurse. https://github.com/CPHSLab/Stress-Detection-in-Nurses (2021). 21.