Early Detection of Chronic Stress Using Wearable Devices: A Machine Learning Approach with the WESAD Database

Amaia Calvo[1] [a], Julen Martin[1] and Cristina Martin[1,2,3] [b]

[1]*Fundación Vicomtech, Basque Research and Technology Alliance (BRTA), Mikeletegi 57, Donostia-San Sebastián, 20009, Basque Coutry, Spain*

[2]*Faculty of Engineering, University of Deusto, Avda. Universidades, 24, Bilbao, 48007, Basque Coutry, Spain*

[3]*BioGipuzkoa Health Research Institute (Bioengineering Area), eHealth Group, 20014 Donostia-San Sebastián, Spain*

Keywords: Chronic Stress, Wearable Devices, Machine Learning, Deep Learning, Stress Detection, Physiological Signals, WESAD Database, Subject-Dependent Models, Health Monitoring.

Abstract: Stress disorders have experienced a significant increase in recent years, impacting individual health. This study explores the feasibility of detecting this mental condition through the analysis of physiological signals captured by wearable devices using machine learning algorithms. An exhaustive review of relevant public databases was conducted and WESAD database was identified as the most suitable one. A detailed examination was conducted using two different configurations for building AI models: in one approach, a single model was created using data from all participants, while in the other, personalized models were developed for each individual participant. This approach evaluated the effectiveness of different preprocessing methods and AI algorithms, as well as identified the physiological signals most informative about stress. Convolutional Neural Networks (CNN) achieved the highest accuracy in stress detection, with an overall accuracy of 99.8% for the single model configuration and 99.6% for personalized models. The analysis also highlighted electrocardiogram (ECG) and electrodermal activity (EDA) as the most informative signals for predicting stress.

1 INTRODUCTION

Stress can be defined as a natural physiological and psychological response activated by situations perceived as threatening or dangerous. It performs an essential role in human alarm and defense mechanisms. Despite this, when these stressful emotions become frequent, they can have harmful effects on mental and physical health, increasing the risk of developing various illnesses, such as cardiovascular diseases, mood disorders, or sleep disorders (Slavich, 2020). Alarmingly, it is estimated that approximately 1 in 4 adults experience stress regularly, and some studies indicate a 30% increase in reported stress levels over the past decade, particularly among younger individuals (American Psychological Association, 2017).

The economic burden of stress-related illnesses on modern societies is substantial, costing healthcare systems billions each year in treatment, lost productivity, and decreased quality of life. According to a report by the World Health Organization (WHO) (Depression, 2017), it is estimated that mental disorders, including stress and anxiety, can cost global economies up to $1 trillion annually in lost productivity. Additionally, a study by Gallup (Gallup, 2017) revealed that employees experiencing stress tend to be less productive, which can negatively impact company profits and the economy as a whole. Work-related stress contributes significantly to productivity loss (Giorgi et al., 2020).

Taking this into account, the early detection of stress becomes crucial to avoid its negative effects (Kivimäki and Steptoe, 2018). Research has emphasized the importance of timely stress detection and the development of preventive solutions to address this growing issue (Slavich, 2020). Furthermore, it has been pointed out that it is essential to create accessible solutions for the entire population to ensure that no one is left without support (Patel et al., 2018). Additionally, studies have indicated that primary care is overwhelmed and that mental health issues continue to rise, highlighting the urgency of implementing effective alternatives (Moise et al., 2021). Traditionally,

[a] https://orcid.org/0009-0009-2806-5344

[b] https://orcid.org/0000-0002-3919-2738

Calvo, A., Martin, J. and Martin, C.
Early Detection of Chronic Stress Using Wearable Devices: A Machine Learning Approach with the WESAD Database.
DOI: 10.5220/0013209700003938

In *Proceedings of the 11th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2025)*, pages 189-196
ISBN: 978-989-758-743-6; ISSN: 2184-4984

stress is measured using self-reported questionnaires or by visiting a mental health practitioner. However, these techniques lack objectivity and are not compatible with everyday situations, highlighting the need for alternative methods to detect stress. Additionally, psychological consultations can be very expensive, making them inaccessible for many people. Stress can develop into more severe mental health conditions if it is not detected in time. Therefore, it is essential to develop affordable, objective, and practical methods for early stress detection to prevent the progression of mental health disorders (Espeleta et al., 2018). The WHO emphasizes that preventive solutions and early interventions are critical components in managing stress and related disorders (World Health Organization, 2018).

In this context, the possibility of detecting stress with wearable devices emerges (Lupton, 2020). Wearables are able to monitor a wide variety of physiological signals such as heart rate, skin temperature and galvanic skin response in an objective, continuous, and affordable way. Furthermore, the recent rise in popularity of these devices, combined with their capacity for continuous monitoring, suggests they could play a key role in future healthcare (Baig et al., 2019) by offering a scalable solution to stress detection that bridges the gap left by traditional methods.

This study aims to investigate the potential for detecting stress using wearable devices, utilizing data from the public WESAD database. A secondary objective of the study is to identify which sensors provide the most critical information for stress detection, enabling the design of experiments ad-hoc to specific requirements for stress prediction in the future.

Through this study, we aim not only to advance the technical performance of stress detection models but also to contribute to the broader goal of developing practical, scalable solutions for early stress detection, with the potential to mitigate the growing burden of stress-related disorders on global health.

The structure of the paper is organized as follows. In Section 2 reviews the related work on physiological signals and notable studies in stress detection. Section 3 describes the WESAD dataset utilized and the preprocessing steps taken. Section 4 outlines the methodology, including the subject-dependent models and preprocessing strategies. Section 5 presents the experiments and results, comparing the performance of various machine learning algorithms. Section 7 discusses the explainability of the models, highlighting the importance of feature relevance. Finally, Section 8 concludes the study by summarizing the key findings and suggesting future research directions.

2 RELATED WORK

2.1 Physiological Signals for Stress Detection

The detection of stress through physiological signals has been extensively studied, leveraging various types of data to assess stress levels accurately. One of the primary physiological indicators of stress is heart rate (HR). Stress typically triggers an increase in HR due to heightened sympathetic nervous system activity. This response is commonly monitored using electrocardiograms (ECG) and photoplethysmograms (PPG). HR measurements are useful in identifying stress, but they need to be complemented by additional metrics for a more comprehensive analysis (Dinh et al., 2020).

Another critical parameter is heart rate variability (HRV), which measures the variation in time between successive heartbeats. HRV is an essential indicator of the autonomic nervous system's responsiveness and adaptability. A reduction in HRV is generally associated with higher stress levels. This measure is derived from the analysis of RR intervals in ECG signals, offering valuable insights into an individual's stress state (Dinh et al., 2020).

Galvanic skin response (GSR), also known as electrodermal activity (EDA), is another widely used physiological signal for stress detection. Stress induces sweating, which changes the skin's electrical conductance. Monitoring GSR can provide significant information about stress levels, especially when used in conjunction with ECG data. However, GSR measurement can be influenced by various factors, including ambient temperature and humidity, which may affect its accuracy (Affanni, 2020) (Eren and Navruz, 2022).

Blood pressure (BP) is also a relevant physiological signal in stress research. Elevated BP can indicate stress, although it may also be influenced by physical exertion and other health conditions. Continuous BP monitoring presents challenges, often requiring indirect measurement techniques such as infrared photoplethysmography (Dinh et al., 2020). While BP data can be informative, it does not always provide a clear distinction between stress-induced and other types of hypertension.

Pupil diameter (PD) has emerged as a promising measure of stress, as stress can cause rapid fluctuations in pupil size. Techniques like video-pupillography are used to measure these changes, but they are often costly and time-consuming, limiting their practical application in real-time stress monitoring (Dinh et al., 2020).

Respiration variability (RESP) is another physiological parameter that reflects stress levels. Stress can alter both the rate and depth of breathing, making RESP measurements valuable for stress assessment. Sensors that track thoracic expansion are used to capture this variability, providing additional data for stress detection (Dinh et al., 2020).

Accelerometers are commonly integrated into wearable devices to monitor involuntary movements such as tremors, which can correlate with stress levels. These devices offer a practical approach to detecting stress-related physical responses and are often used in combination with other physiological measures to enhance detection accuracy (Dinh et al., 2020).

2.2 Notable Studies

Several studies have made substantial contributions to the field. For instance, a study published by IEEE in 2012 achieved an 81% accuracy rate in distinguishing between stressed and non-stressed states using a wearable device that measured ECG, GSR, electromyography (EMG), and respiratory frequency (Can et al., 2020). This research focused on detecting acute stress rather than chronic stress, inducing stress in participants through psychophysiological tasks designed to elicit specific mental states. The study involved a relatively small sample of 20 participants, including both men and women, monitored continuously for over 13 hours. The logistic regression model used for classification demonstrated the potential of wearable devices for continuous stress detection, although the accuracy suggests a need for more sophisticated models to enhance the precision.

Another notable study conducted by Bogazici University and the University of Milan in 2020 achieved a 94.52% accuracy rate in classifying stress levels using a hybrid artificial intelligence approach (Can et al., 2020). This study also targeted acute stress but included a wider range of stress levels, differentiating between low, moderate, and high stress. Stress was induced during a structured event comprising baseline, lecture, exam, and recovery sessions, allowing researchers to analyze how stress management techniques, specifically guided mindfulness, affected stress levels. The sample consisted of 32 participants, with demographic details not extensively discussed. The dataset was collected across various sessions, enhancing the practical applicability of the findings. The use of everyday wearable devices such as smartwatches allowed for unobtrusive and continuous monitoring, improving accuracy through personalized stress clustering and decision-level smoothing techniques to correct misclassifications.

Additionally, research from the University of Vigo investigated wearable devices for stress and sleep monitoring, achieving a 90% accuracy rate with various machine learning models (Dalmeida and Masala, 2021). This study focused on acute stress experienced by 27 young, healthy participants while driving. The dataset included physiological signals measured during different driving conditions—rest, highway, and city driving—utilizing physiological signals to develop predictive models. Multiple machine learning algorithms, including K-Nearest Neighbor (KNN) and Support Vector Machines (SVM), were tested, with SVM achieving the highest performance at 83.33% accuracy. While this study provided insights into real-world stress detection in driving scenarios, it faced challenges typical of real-life applications, such as variations in accuracy compared to laboratory settings.

Despite these advancements, several challenges remain in the field of stress detection using physiological signals (Dalmeida and Masala, 2021). One significant limitation is the precision and sensitivity of wearable devices, which can vary widely and be influenced by factors unrelated to stress. Additionally, the cost of high-quality wearables can be prohibitive, limiting their accessibility compared to clinical devices. Finally, the specificity of different devices and measurement techniques can lead to inconsistencies in stress detection results, highlighting the need for standardized approaches.

This review underscores the significant progress made in physiological signal analysis and machine learning for stress detection. While advancements continue to enhance the accuracy and practicality of stress monitoring, ongoing research is needed to address current limitations, such as generalizability across different populations and the integration of multi-modal physiological signals. Additionally, while accuracy is an important metric, it should not be the sole focus; other performance indicators like recall and precision are essential for evaluating model robustness in real-world applications.

3 DESCRIPTION OF THE DATASET

The dataset chosen for this study is the publicly available WESAD (Wearable Affect and Stress Detection) dataset, designed for the analysis of acute stress responses rather than chronic stress. This dataset was selected for its rich physiological data, making it suitable for studying short-term stress detection through wearable devices.

The dataset includes data from 17 volunteers who underwent stress induction procedures in a controlled laboratory. After excluding 2 subjects due to data interference, the final dataset contains 15 participants: 12 men and 3 women, with an average age of 27.4 years.

It records physiological signals across three emotional states: baseline, stress, and amusement. The baseline phase lasted 20 minutes, followed by a 10-minute Trier Social Stress Test (TSST) to induce stress, and finally, 6 minutes of comical videos to elicit amusement. Although amusement data is available, it is not used in this study, which focuses solely on stress detection.

Each participant has approximately 36 minutes of data. Data was collected using the RespiBAN Professional chest band and the Empatica E4 smartband. The RespiBAN captures higher-quality data with a 700 Hz sampling rate for respiratory rate, accelerometer, ECG, EDA, EMG, and temperature. The Empatica E4, with lower sampling rates, recorded blood pressure, EDA, temperature, and accelerometer data. Due to numerous missing values in the Empatica E4 signals, only data from the RespiBAN is included in this study.

4 METHODOLOGY

This study verifies the possibility of predicting stressful emotional states using physiological signals from two different configurations of **subject dependent models**. The objective is to learn how physiological signals of a person can be used to detect her/his stress level.

4.1 Subject-Dependent Models

A subject-dependent approach utilizes data from the same individual for training, validation, and testing phases of model creation. One advantage of this strategy is that it allows the model to become more personalized by learning the unique characteristics of each person. On the contrary, when trying to identify the stress of another (different) individual they might not generalize well.

In this study, the data was divided into a training, validation and test subset while maintaining the temporal structure of the signals data. The training subset is composed of the first 70% of the data, the validation subset consisted of the next 15%, and the test subset included the final 15%.

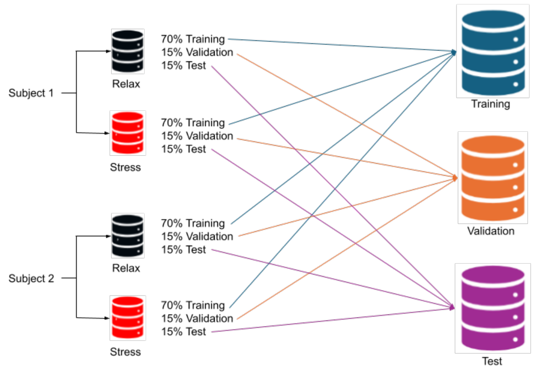

Figure 1: Data partitioning into subject-dependent models with all participants.

Subject-dependent models were further divided into two configurations:

- **One Single Model for all Participants:** In this configuration, data from all participants are used in training, validation, and testing phases. This results in a model that attempts to generalize across multiple individuals while maintaining the temporal structure of the signals (see Figure 1).
- **One Personalized Model per Participant:** In this configuration, a separate model is trained for each individual participant. This allows for a highly personalized approach, where the model only learns from the specific individual's data. The same 70-15-15 temporal split is used for training, validation, and testing, but exclusively with the data from one subject at a time (see Figure 2).

The main goal of this comparison is to assess whether using data from multiple participants enhances model performance by providing a wider range of variability, or if personalized models that focus on individual patterns yield better predictive accuracy due to their specificity to one subject.

4.2 Preprocessing Strategies

In addition to optimize the performance of the subject-dependent models, this study will evaluate the effect of three different preprocessing strategies:

- **P1:** Applies Min-Max normalization, which scales the data into the range [0, 1]. This process help mitigate the impact of features with larger values by scaling all variables to a common range.
- **P2:** Normalizes the data similarly to P1, but followed by SMOTE (Synthetic Minority Oversampling Technique) to address class imbalance by

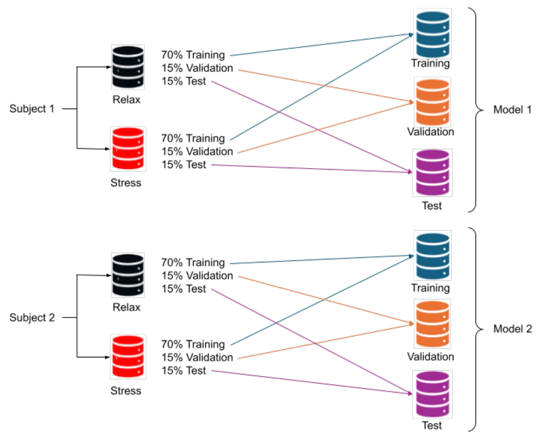

Figure 2: Data partitioning for personalized models.

generating synthetic examples for the minority class.

- **P3:**This technique initially applies normalization to the data, followed by the implementation of PCA (Principal Component Analysis) to reduce the dimensionality of the data and capture the most important features while discarding redundant information.

4.3 Evaluation Metrics

The evaluation of the models was performed using two key metrics: **accuracy** and **F1-score**.

- **Accuracy** is defined as the proportion of correctly classified cases out of the total. It is commonly used when all classes are equally important.
- **F1-score** is the harmonic mean of precision and recall, and is especially useful when minimizing false negatives is critical, such as in medical applications. Precision represents the proportion of true positive predictions out of all positive predictions, while recall (or sensitivity) represents the proportion of true positives out of the actual positive cases.

The results for each combination of model and preprocessing technique will be presented below.

5 EXPERIMENTS AND RESULTS

5.1 Algorithms Used

Machine learning includes a wide range of algorithms for classifying new instances. This study aims to compare the effectiveness of several algorithms. Among the machine learning algorithms that will be evaluated are Decision Trees, Random Forests, Support Vector Machines (SVM), Adaboost, Logistic Regression, XGBoost, Linear Discriminant Analysis (LDA), and K-Nearest Neighbours. In addition, the performance of deep learning algorithms, such as LSTM Recurrent Neural Networks and Convolutional Neural Networks (CNN), see Table 1.

5.2 Results

The goal is to provide a comparative analysis to identify which algorithms offer the best performance in terms of accuracy and F1-score for this dataset. Table 1 allows a better selection of the optimal algorithm to make more accurate predictions about the stress state of patients. It is worth noting that the study focuses on binary classification (stress and relaxation).

The AI models trained are used for the two different model configurations and provide the following metrics:

- **One Single Model for all Participants:** The best-performing algorithm was the Convolutional Neural Networks (CNN). CNN achieved the highest performance in binary classification, with an accuracy of 99.8% and F1-score of 0.998. The results compared with P1 indicate that CNN not only provides the best option in terms of accuracy and F1 of the deep learning algorithms but also highlights the best option among the machine learning algorithms.
- **One Personalized Model per Participant:** Personalized models were built for each of the 15 subjects, and the average performance across these models was calculated. Once again, CNN proved to be the top performer in binary classification, achieving an accuracy of 96.4% and an F1-score of 0.962. These results indicate that CNN not only stands out as the best option in terms of accuracy and F1 for personalized models but also reinforces its position as the leading choice among machine learning algorithms.

In terms of overall performance, models trained with data from all subjects (general models) tended to outperform personalized models. This suggests that a more diverse dataset improves the model's ability to generalize and classify new instances more effectively. While CNN performed exceptionally well (An accuracy of 99.8% with P1 and an F1-score of 0.998.), the results indicate that tree-based models like Random Forest may not yield the same level of accuracy in this binary context. Moreover, deep learning models such as CNN appeared to benefit more from personalization and appropriate preprocessing, reinforc-

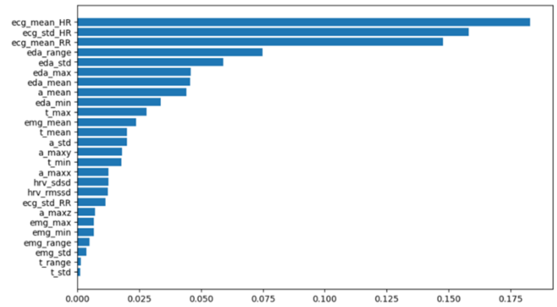

Figure 3: Feature Importance.

ing the notion that an individualized approach can be advantageous when tailored correctly.

6 EXPLAINABILITY

The analysis of explainability is conducted for the general model designed using Convolutional Neural Networks (CNN). As the integration of machine learning applications in society increases, the explainability of predictive models is becoming an essential aspect. Explainability provides transparency in model decisions, which is crucial in the field of medicine. In this context, explainability can help identify the most relevant variables for predicting stress, thereby enabling the design of more effective and personalized interventions for its management and reduction. To evaluate the explainability of the obtained model, the importance of features and SHapley Additive exPlanations (SHAP) values are assessed.

Feature importance assigns a score to each feature, indicating its relevance in model construction. Features with higher scores are considered more important. The results of the feature importance analysis are presented below. Figure 3 illustrates that the most important features are derived from electrocardiogram and electrodermal activity sensors. This provides insight into which variables are most affected by stressful situations, indicating which sensors are most useful as biomarkers.

On the other hand, SHAP values offer a method for explaining a predictive model's response based on game theory. They measure how much each variable contributes to the prediction of a given observation, allowing for a more detailed and precise interpretation of how individual features affect predictions. One advantage of SHAP values is that they indicate whether each variable has a positive or negative impact on predictions based on its values. Another benefit is that SHAP values enable local interpretability; that is, one can arbitrarily select an instance to examine which factors were most relevant in predicting that specific case.

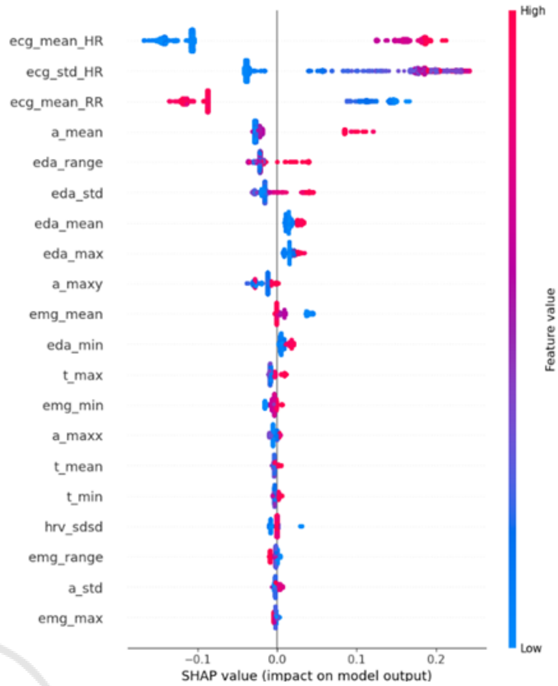

Figure 4: SHAP Values.

Figure 4 displays the SHAP values for all predictions from the last fold of cross-validation. A high SHAP value indicates that the variable significantly impacts the model's prediction, while values close to 0 reveal that the variable has little influence on the results. Analyzing this image shows that the most influential variables are the mean and standard deviation of heart rate, the mean of accelerometer readings, and the mean, range, maximum, and standard deviation of electrodermal activity (EDA). Notably, while both methods indicate the same five main variables for prediction, the feature importance analysis does not include the mean of accelerometer readings (a_mean), which does appear in the SHAP values. This discrepancy may reflect differences in how each method evaluates feature relevance.

Considering the results from both explainability techniques, it can be concluded that the most relevant sensors for stress detection are the ECG, EDA, and accelerometer.

7 RESULTS DISCUSSION

After analyzing stress detection using the WESAD dataset, several significant conclusions were drawn. The initial step involved exploring the data to understand its distribution and the potential relevance of each sensor in the prediction. Subsequently, the time

Table 1: Comparison of Classification Models.

	Subject-dependent models					
	P1		P2		P3	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Decision trees	0.922	0.922	0.934	0.933	0.935	0.935
Random Forest	0.993	0.993	0.996	0.996	0.996	0.996
SVM	0.973	0.973	0.973	0.973	0.974	0.974
Adaboost	0.990	0.990	0.987	0.987	0.919	0.919
Logistic Regression	0.893	0.892	0.885	0.885	0.859	0.859
XGBoost	0.949	0.949	0.943	0.943	0.963	0.963
LDA	0.878	0.876	0.849	0.849	0.865	0.862
KNN	0.973	0.973	0.973	0.973	0.993	0.993
LSTM	0.965	0.965	0.953	0.953	0.945	0.945
CNN	**0.998**	**0.998**	0.983	0.983	0.979	0.979
	Personalized models					
	P1		P2		P3	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Decision trees	0.899	0.889	0.760	0.708	0.944	0.933
Random Forest	0.985	0.985	0.978	0.978	0.906	0.958
SVM	0.979	0.980	0.959	0.957	0.962	0.962
Adaboost	0.932	0.933	0.760	0.708	0.958	0.949
Logistic Regression	0.977	0.975	0.882	0.875	0.945	0.935
XGBoost	0.912	0.898	0.971	0.971	0.916	0.903
LDA	0.978	0.979	0.959	0.959	0.956	0.954
KNN	0.965	0.962	0.956	0.956	0.963	0.960
LSTM	0.916	0.911	0.946	0.941	0.925	0.921
CNN	0.964	0.962	0.949	0.947	**0.996**	**0.996**

series data was preprocessed and transformed into a tabular format using the sliding window technique, which facilitated the extraction of features.

The data was then divided into subject-dependent configurations, and various machine learning algorithms were applied to determine the most effective one. Regarding the initial hypothesis of the project, it can be stated that it is feasible to develop a stress prediction model using information collected from wearable devices. After analyzing several machine learning algorithms, the one offering the best results was selected for both subject-dependent models. In the subject-dependent models, the Convolutional Neural Network (CNN) achieved an accuracy of 99.8% for binary classification.

The analysis of subject-dependent configurations revealed that a dataset with more users generally yields better results than personalized models. This indicates that, despite belonging to different subjects, the inclusion of a larger volume of data provides generalizable information, improving the accuracy of predictions. Notably, these general models can be likened to a laboratory setting, where diverse participants contribute to a richer dataset, similar to a football team training together. In contrast, personalized models—tailored for individual subjects—provide a more precise approach, akin to customized care plans in primary care settings. This understanding of individual variability enhances the effectiveness of interventions.

Additionally, it was observed that the treatment of data imbalance did not significantly influence the results as anticipated; in many cases, the best-performing model was the one that did not apply SMOTE. In the subject-dependent models, the importance of selecting the appropriate window size for feature extraction was highlighted. The best results were obtained with 4-minute windows and a step size of 1 second, which are relatively wide for this type of case. This allows for a more comprehensive view of the time series, capturing the global characteristics of each class and reducing signal noise. However, a large window size may overlook important details in different types of signals and can increase the computational complexity of processing. The optimization phase also concluded that, despite having very different characteristics, a higher number of subjects in the study improves model performance.

Finally, the study of explainability provided insights into the relevance of each variable in the model. Two methods were employed: feature importance and SHAP values. The results from both methods indicated that the most impactful variables for prediction were those derived from the ECG, accelerometer, and EDA sensors.

8 CONCLUSIONS

This study confirms the feasibility of developing an effective stress prediction model using information collected from wearable devices. The findings underscore the importance of leveraging diverse datasets to enhance predictive accuracy, as demonstrated by the Convolutional Neural Network achieving an accuracy of 99.8% for a binary classification that identifies relaxed-stressed situations using two subject dependent models. By building AI models upon WESAD dataset we learned that ECG and EDA signals provide the most valuable information to predict stress. The results obtained in this research work will be used in an observational study that will build a new dataset to predict the stress suffered by Vicomtech professionals.

The potential applications of these models extend to real-world settings, where early stress detection can lead to timely interventions and improved mental health outcomes. Future research could focus on optimizing personalized models for individual subjects and exploring the integration of additional physiological data from commercial wearables to advance early stress detection.

ACKNOWLEDGMENT

We would like to acknowledge *QOLIFE - Multimodal real world capture and processing for quality of life assessment* project, funded as internal project of Fundación Vicomtech.

REFERENCES

Affanni, A. (2020). Wireless sensors system for stress detection by means of ecg and eda acquisition. *Sensors*, 20(7):2026.

American Psychological Association (2017). Stress in america: The state of our nation. Technical report, American Psychological Association.

Baig, M. M., Afifi, S., GholamHosseini, H., and Mirza, F. (2019). A systematic review of wearable sensors and iot-based monitoring applications for older adults–a focus on ageing population and independent living. *Journal of medical systems*, 43:1–11.

Can, Y. S., Chalabianloo, N., Ekiz, D., Fernandez-Alvarez, J., Riva, G., and Ersoy, C. (2020). Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches. *IEEE Access*, 8:38146–38163.

Dalmeida, K. M. and Masala, G. L. (2021). Hrv features as viable physiological markers for stress detection using wearable devices. *Sensors*, 21(8):2873.

Depression, W. (2017). Other common mental disorders: global health estimates. *Geneva: World Health Organization*, 24(1).

Dinh, T., Nguyen, T., Phan, H.-P., Nguyen, N.-T., Dao, D. V., and Bell, J. (2020). Stretchable respiration sensors: Advanced designs and multifunctional platforms for wearable physiological monitoring. *Biosensors and Bioelectronics*, 166:112460.

Eren, E. and Navruz, T. S. (2022). Stress detection with deep learning using bvp and eda signals. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–7. IEEE.

Espeleta, H. C., Brett, E. I., Ridings, L. E., Leavens, E. L., and Mullins, L. L. (2018). Childhood adversity and adult health-risk behaviors: Examining the roles of emotion dysregulation and urgency. *Child Abuse & Neglect*, 82:92–101.

Gallup (2017). State of the american workplace: Employee engagement insights for u.s. business leaders. Technical report, Gallup.

Giorgi, G., Lecca, L. I., Alessio, F., Finstad, G. L., Bondanini, G., Lulli, L. G., Arcangeli, G., and Mucci, N. (2020). Covid-19-related mental health effects in the workplace: a narrative review. *International journal of environmental research and public health*, 17(21):7857.

Kivimäki, M. and Steptoe, A. (2018). Effects of stress on the development and progression of cardiovascular disease. *Nature Reviews Cardiology*, 15(4):215–229.

Lupton, D. (2020). Wearable devices: Sociotechnical imaginaries and agential capacities.

Moise, N., Wainberg, M., and Shah, R. N. (2021). Primary care and mental health: Where do we go from here? *World Journal of Psychiatry*, 11(7):271–276.

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., et al. (2018). The lancet commission on global mental health and sustainable development. *The lancet*, 392(10157):1553–1598.

Slavich, G. M. (2020). Social safety theory: a biologically based evolutionary perspective on life stress, health, and behavior. *Annual review of clinical psychology*, 16(1):265–295.

World Health Organization (2018). Mental health: strengthening our response. Technical report, World Health Organization.