This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

Contribution of EEG Signals for Students' Stress Detection

Jonah Fernandez, Raquel Martínez, Bianca Innocenti, Beatriz López

Abstract—Stress is a prevalent global concern impacting individuals across various life aspects. This paper investigates stress detection using electroencephalographic (EEG) signals, which have proven valuable for studying neural correlates of stress. Stress was induced in students, and physiological data was recorded as part of the experimental setup. Different feature sets were extracted and four machine learning models, including LightGBM, Convolutional Neural Network (CNN), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), were utilized for classification tasks. The findings indicate that the mean and standard deviation of 19 channels consistently outperform other feature sets. LightGBM demonstrates superior performance across all scenarios compared to CNN, KNN, and SVM. Overall, this study presents an effective stress detection approach using EEG signals and demonstrates the potential of integrating simple statistical features for enhanced classification accuracy. The findings contribute to the advancement of stress monitoring technologies, with potential applications in wearables and BCIs for real-time stress management.

Index Terms—Stress, Electroencephalogram (EEG), emotion recognition, feature extraction, feature selection, machine learning

1 Introduction

Stress is a natural human response to challenges and threats, influenced by both external factors like living conditions and internal factors such as cognitive processes [1], [2]. When stress becomes chronic and goes undetected, it can lead to serious mental health issues [3]. Effective stress detection is crucial for preventing these conditions. While psychological methods have traditionally been important for measuring stress, recent studies emphasize the growing importance of using physiological indicators for a more accurate assessment [4] [5]. This work focuses on developing stress detection solutions using biosignals, with particular emphasis on the contribution of EEG (Electroencephalography) signals. Traditional methods rely on manual examinations by health professionals, which are time-consuming and expensive [4]. This research aims to automate stress detection through physiological data, reducing the burden on both patients and healthcare systems.

Among various physiological measures, EEG signals stand out due to their ability to provide a direct reflection of neural activity during stress conditions [6]. While other measures like electrocardiogram (ECG) [7], electrodermal activity (EDA) [8], and respiration rate (RSP) [9] also show consistent stress-related changes, EEG offers unique insights into how the brain responds to stressors by capturing electrical patterns that correlate with cognitive and emotional processes [10]. This makes EEG devices particularly valuable for understanding the complex dynamics of stress at the neural level.

Furthermore, the application of machine learning methods to biosignal data enables the automatic detection of stress states. For example, Heyat et al. [11] used Decision Tree (DT), Naive Bayes (NB), Random Forest (RF), and Logistic Regression (LR) to classify subjects in mental stress and non-stress. The study was carried out considering the data available for a single subject (intra-subject classification) as well as taking into account all available data (inter-subject classification). Similarly, Hosseini et al. [12] focused on stress detection using normalized EDA signals and employing machine learning models such as AdaBoost, RF, and Support Vector Machines (SVM). Asif et al. [13] examined the effect of music tracks on human stress levels using EEG signals from twenty-seven subjects. The authors of [13] used sequential minimal optimization, stochastic gradient descent, LR, and Multilayer Perceptron (MLP) to classify the subject's stress level into two (stressed and not-stressed) and three classes (non-stressed, medium stressed and highly stressed). However, the application of machine learning to biosignal data requires a preliminary signal processing step [14] Biosignal data comes in the form of times series sampled at a given frequency, with eventually some noise and artifacts. After a cleaning phase, a common practice is to reduce the original data to the lowest dimensional space by extracting features that resume them. For example, Heyat et al. [11] calculated the following features from ECG records: the average heart rate (AHR) and mean of the RR interval (distance between signal peaks). In the case of the EEG signals, features in the time domain (as principal component analysis, independent component analysis), the frequency domain (i.e. Fourier transform, power spectral density (PSD)), and time-frequency domain (as spectrogram-based features or wavelet analysis-based features) are considered [15]. While feature extraction reduces the computational burden for subsequent processing, the selection of an optimal feature set remains challenging, as each method has inherent limitations [10]. For instance, PSD interpretation requires domain expertise, while wavelet analysis depends on the choice of an appropriate function to balance time and frequency resolution. Moreover, despite the computational benefits in later stages, the process of computing these features itself can be resource-intensive, adding to the overall computational cost.

Additionally, with the advances in sensor and wearable technologies, the use of biosignals outside of the clinical environment could provide an invaluable source of information to understand patterns of stress in daily life contexts. Regarding EEG, simple devices that do not require conductive gels and other complex installations to measure brain activity, can be used for monitoring health conditions [16]. This kind of device, however, provides a single biosignal. The question is if the EEG data gathered by simple biosensors could be enough regarding stress detection, or should be complemented by other devices, such as wristbands that provide additional physiological data. Most of the studies analyzing EEG focus on machine learning performance while ignoring the analysis of how EEG contributes to stress detection regarding other biosignals.

In this context, our study serves as a crucial bridge, aiming to evaluate the substantial contribution of EEG to stress prediction when compared to established physiological measures such as ECG and EDA. In doing so, we analyze which EEG channel could provide more relevant data regarding stress detection, to condition the purchase of a simple EEG wearable for such purpose. As a secondary contribution, we explore several feature extraction methods, revealing that simple, low-cost computational ones, could be enough for detecting stress.

2 Related Work

Recent research has introduced innovative methodologies that utilize biosignals, with a particular emphasis on EEG as the primary signal of interest, to detect and classify stress at different levels. Samarpita et al. [17] trained a diverse set of machine learning algorithms, such as RF, DT, K-Nearest Neighbors (KNN), MLP, SVM, Adaboost, and Extreme Gradient Boosting (XGBoost). Baliga et al. [18] extracted only the alpha (8-13 Hz) and beta (13-30 Hz) bands from EEG records and then applied two binary classification algorithms (”under stress” or ”not under stress”), SVM and LSTM, conversely to our work, which considered the full spectrum of EEG signals for a more comprehensive analysis. Marthinsen et al. [19] proposed a cost-effective minimally intrusive framework with only eight EEG channels selected with a Genetic Algorithm (GA). Features were extracted using Wavelet Scattering and they employed two classification algorithms, SVM and Convolutional Neural Network (CNN), for their analysis. These works differ from ours in several key respects. While [17] focused solely on 4 subjects and 4 channels, we leveraged data from 30 participants across 19 channels. [18] employed data from 40 subjects, but with recordings of only 80 seconds duration, whereas our recordings extend to approximately 22 minutes. In addition, the increased number of channels and longer recording duration may contribute to a richer feature space and potentially enhance spatial resolution in our study compared to [19]'s 8-channel selected approach.

Alternatively, Alshorman et al. [20] applied Fast Fourier Transform (FFT) as a feature extraction stage to measure all bands' power density for the frontal lobe and then used SVM and NB classifiers. Similarly, Nirabi et al. [21] presented a feature extraction method using discrete wavelet transform (DWT). The features were classified using KNN, SVM, NB, and LDA. Our study takes a distinct approach compared to previous works by [20] and [21]. While they relied on complex feature extraction methods like FFT or DWT, this work utilizes simpler statistical features based on means and standard deviations.

Other studies in stress detection have focused on the combination of multiple biosignals. Hemakom et al. [22] used EEG and ECG signals, and Zontone et al. [23] proposed a method that combines EDA and ECG signal analysis to detect stress in car drivers. Similarly, Affani [24] introduced a stress detection design with two EDA sensors and two ECG sensors, exhibiting high-level performances in terms of linearity and jitter during metrological characterization. Our approach differs significantly from some of these works in the number of features employed. While we utilize 38 features, the methodologies of other studies include a substantially smaller set: 5 features in [20] and [24], 9 in [23], and 10 in [21]. Notably, [22] utilized a much larger feature set (147 features) derived from only 8 EEG channels. While a larger number of features can potentially capture a wider range of information, it also increases the risk of including redundant or irrelevant features. These redundant features can lead to several issues, including increased computation time, overfitting, model complexity, and difficulty in feature selection.

On the other hand, some studies have focused on using multimodal data for stress detection. For instance, Umematsu et al. [25] examined how multi-modal data, including physiological signals (EDA), mobile phone usage, location, and behavioral surveys collected over the previous N days, can forecast tomorrow evening's stress levels in students. Alternatively, Bobade and Vani [26] used multimodal data extracted from wearable physiological and motion sensors, including three-axis acceleration (ACC), ECG, blood volume pulse (BVP), body temperature (TEMP), respiration (RESP), electromyogram (EMG), and EDA. While [25] and [26] employed multimodal data excluding EEG, our study focused on EEG data and its contribution to stress detection. This approach avoids the complexities and potential confounding effects associated with integrating multiple modalities, streamlining the stress detection process.

Table 1 presents the summary of the related work section. Unlike previous studies, this work uniquely incorporates a channel importance analysis for EEG data. This analysis can pave the way for the development of more focused and efficient stress-detection wearables in the future. By identifying the most informative EEG channels, we can potentially reduce the number of electrodes needed in wearable devices, leading to more comfortable and user-friendly designs. Additionally, we conducted a comparative performance study of EEG against ECG and EDA, examining the effectiveness of these signal types in stress detection applications. Our review of the literature also revealed a prevalent absence of LightGBM models for classification, despite their potential to outperform traditional methods, particularly in handling large datasets with greater speed. Moreover, our approach leverages a low computational cost due to relatively simple feature extraction techniques, ensuring the efficient extraction of relevant information from these signals.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

Ref.	Data	# Subjects	# EEG channels	EEG Features	# Features	# Labels	EEG and ECG/EDA comparison	EEG channel importance study
Samarpita et al. [17]	EEG	4	4	PSD (3 bands)	Unspecified	3	No	No
Baliga et al. [18]	EEG	40	32	DWT (2 bands)	Unspecified	2	No	No
Marthinsen et al. [19]	EEG	28	8	Wavelet scattering	Unspecified	2	No	No
Alshorman et al. [20]	EEG	14	128	FFT (5 bands)	5	2	No	No
Nirabi et al. [21]	EEG	25	7	DWT (5 bands)	10	2	No	No
Hemakom et al. [22]	EEG, ECG	40	8	FFT (7 bands)	147	3	Yes (only EEG and ECG)	No
Zontone et al. [23]	EDA, ECG	16	0	-	9	2	No	-
Affani [24]	EDA, ECG	10	0	-	5	2	No	-
Umematsu et al. [25]	EDA, Other	142	0	-	375	2	No	-
Bobade and Vani [26]	EDA, ECG, Other	15	0	-	39	2, 3	No	-
This work	EEG	30	19	Mean and std	38	3	Yes	Yes

TABLE 1: Related work summary.

3 Methods

In this section, we presented the fundamental framework of the proposed stress recognition system (see Fig. 1). Our aim is to analyze the contribution of the EEG biosignal. To achieve this, two main pathways are defined in the methodology: one that considers EEG data and another one that deals with ECG and EDA data. Therefore, data collection involves obtaining EEG, ECG, and EDA signals from the same subject, using the same protocol, to enable the comparison of the results. Subsequently, signal processing techniques are applied to clean the data and ensure consistency across all signals. Next, several feature extraction methods are applied to transform the longitudinal nature of the biosignals (time series data) into a tabular representation. Following this, the data is labeled (stress, neutral, and relax) and balanced across different stress classes. Finally, to evaluate their efficacy in stress detection, various machine learning methods are employed, leading to the building of different detection models.

3.1 Data Collection Protocol

A total of 30 subjects (fifteen men and fifteen women) were recruited for this study. All participants were students, ranging in age from 18 to 38 years (mean = 20.8, SD = 3.4). The selection of an equal number of male and female participants aimed to mitigate potential gender biases in emotional responses. This study had already been approved by the corresponding ethical committee and met all the criteria required by the current regulations (CEISH-UPV/EHU, BOPV 32 (M10_2016_189)).

3.1.1 Devices Description

The acquisition of the participants' data was carried out with the EEG Enobio device from Neuroelectrics and the BIOPAC MP36 hardware (Biopac Systems Inc., USA) for physiological signals. For the acquisition of EEG signals using the Enobio device, a configuration with 19 standard electrodes was employed, following the setup illustrated in Fig. 2. The data acquisition process was facilitated by the *NIC2* software, ensuring a sampling rate of 500 Hz.

BIOPAC has one set of electrodes for capturing ECG signals and another for EDA signals. The EDA sensors were placed on two fingers of the subjects. For ECG signals, the electrodes were placed following the configuration shown in Fig. 3. The signals were processed using the *BIOPAC Student Lab* software, employing a three-channel configuration. Channel 1 was designated for recording ECG signals and Channel 2 for EDA signals. The sampling frequency for all three channels was uniformly set at 500 Hz.

Fig. 4 illustrates the comprehensive configuration of electrodes used in the study to capture EEG, ECG, and EDA signals. The image provides a visual representation of the electrode placement on the subject. This electrode configuration was employed during data collection to ensure accurate and simultaneous acquisition of multiple physiological signals, allowing for a holistic analysis of the subject's stress responses.

3.1.2 Tests Description

The experimental procedure adopted in this study was based on the approach previously developed by Martinez and collaborators [5], who conducted two experiments in a group of participants, one of them seeking to induce stress. The experiment consisted of a single activity, with a pre and post-phase taking the subject to a basal state of relaxation by watching a relaxing video. As emotional responses to stimuli can vary significantly between individuals, the use of a single activity was concluded to be poor in capturing the subjects' variability.

Therefore, in this study, we chose to expose participants to a variety of stressful stimuli, to address a broader range of stressful situations. To achieve this purpose, the experimental procedure encompassed three stress induction tests, interspersed with intervals during which subjects were presented with two-minute relaxation videos. These stress induction tests comprised the use of 3D puzzles [5], mathematical calculations widely used in scientific literature [28] [29] and, finally, the implementation of a video game on mobile devices. Apart from the tests themselves, an attempt was made to generate a stressful environment and the subject's reactions and behaviors were noted while monitoring the experiment.

Fig. 5 provides a comprehensive view of the complete test sequence employed in the study. The details of every step are the following:

- **3D Puzzle:** In this test, the subject was asked to solve a 3-dimensional puzzle in a limited time. After watching a relaxing video, the subject had a limit of seven minutes to solve the puzzle. When there were only two minutes left, the subject was alerted to try to make them nervous. Subsequent to the test, the subject was presented with the solution, whereby it

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

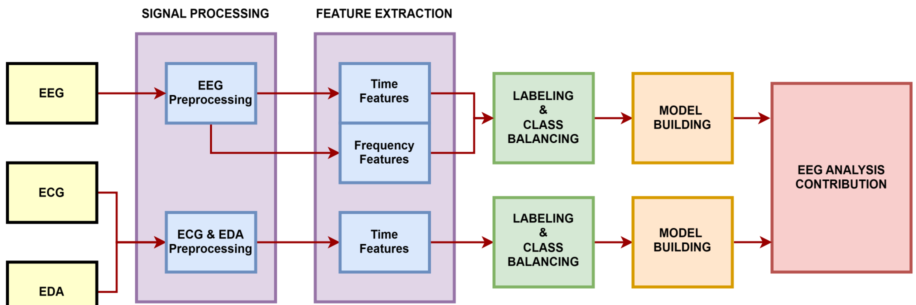

Fig. 1: Flowchart of the proposed method.

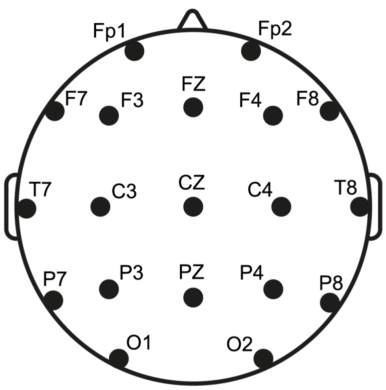

Fig. 2: EEG standard electrode configuration

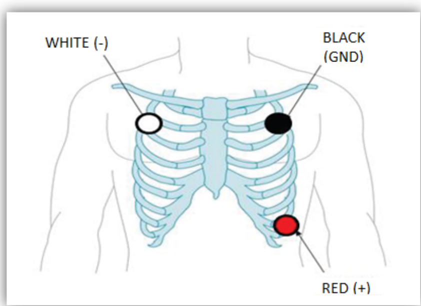

Fig. 3: ECG Electrode Positioning

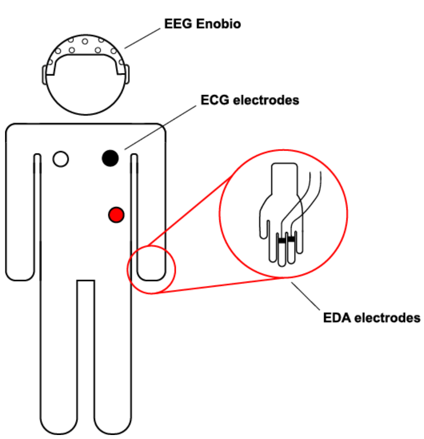

Fig. 4: Electrode configuration (EEG, ECG and EDA)

was occasionally observed that comprehending the solution on paper proved more challenging than engaging in the puzzle itself. This discrepancy between the perceived difficulty of solving the puzzle and the actual complexity of understanding the solution introduced an additional source of frustration and stress.

- **Mathematical Calculations:** In this test, the experimenter sequentially proposed mathematical calculations of increasing difficulty. As the subject answered each question correctly, another calculation of greater difficulty was proposed. The objective was to answer the maximum number of calculations in a limited time, in this case, three minutes.
- **Video Game:** In this experiment, the subject was provided with a musical mobile game that incorporates increasing difficulty. Depending on the subject's knowledge of this type of video game, the difficulty was adjusted to challenge the subject.

At the end of the experiment, an interview and a questionnaire were conducted to contrast this information with the data gathered by the sensors. Post-experiment interviews with the subjects provided valuable insights into the participants' subjective experiences during specific tests and moments throughout the experiment. Participants were encouraged to express their feelings and perceptions, allowing researchers to contrast this qualitative information with the physiological data.

3.2 Signal Processing

Signal processing techniques were applied to the raw data (see Fig. 6) to ensure the integrity and reliability of the subsequent analysis [30].

EEG preprocessing. The initial step involved the application of a band-pass filter ranging from 0.1 to 45 Hz, enabling the identification of electrodes with noise or no discernible signal. Such electrodes were flagged as "bad" in the initial configuration, facilitating their exclusion from further analysis.

To address the presence of ocular artifacts, an Independent Component Analysis (ICA) was employed. ICA is a widely accepted approach that aims to identify linear transformations minimizing the statistical dependence between components in the signal. This enables the isolation and removal of artifactual sources, particularly those associated

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

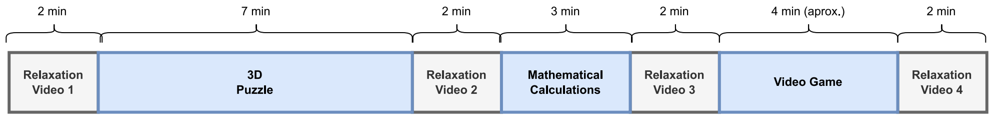

Fig. 5: Tests sequence

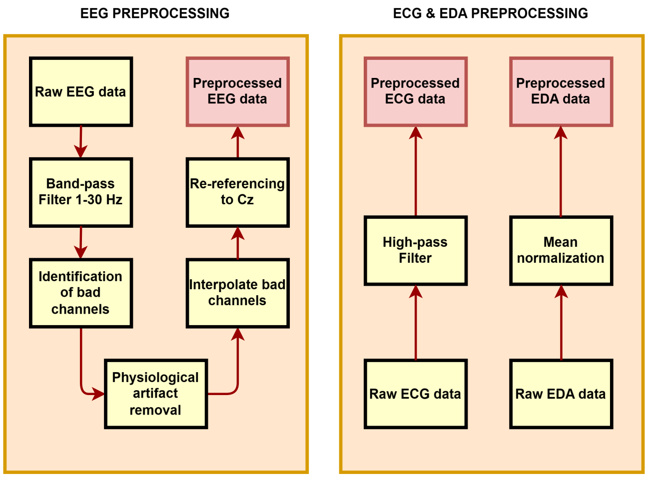

Fig. 6: Signal Processing

with eye movements and blinks. By effectively eliminating these artifacts, the EEG signal can be corrected, preventing the loss of valuable information essential for accurately detecting the subject's emotional state [31].

Given the multi-subject nature of the analysis, it was imperative to maintain consistent dimensionality across all subjects. To achieve this, interpolation of the bad channels was preferred over their outright removal, as the set of bad channels may not coincide between subjects. The spherical spline method was employed for channel interpolation [32], involving the projection of sensor locations onto a unit sphere. Subsequently, the signal at the bad sensor locations was interpolated based on the signals obtained from the surrounding good locations.

In the final preprocessing step, a re-referencing procedure was implemented to ensure a consistent reference point across all channels. The earlobe (GND) reference was transformed to the Cz electrode, chosen for its symmetrical configuration. This re-referencing process further standardized the EEG data, enhancing the comparability and coherence of subsequent analyses.

ECG and EDA preprocessing. A high-pass filter was applied to eliminate noise, enhancing the integrity of the ECG data. In the case of EDA signals, mean normalization was employed, considering that each subject may exhibit varying levels of sweating, and environmental conditions such as temperature and humidity in the laboratory may not be uniform across subjects.

3.3 Feature Extraction

Once the signal is cleaned, in the feature extraction step the longitudinal data captured by the signals is transformed into a tabular approach, so obtaining a set of features per subject. To that end, first, the biosignal is split into windows, and then the information of the windows is summarized in different features.

EEG. The total recording duration was approximately 22 minutes. EEG data were recorded at 500Hz, and windows of 500 samples were selected to represent 1 second of EEG data, considering the sampling frequency of 500 samples per second. Each window is an instance for further analysis.

Then, two distinct approaches were employed for feature extraction.

The first approach revolved around incorporating time domain features. We opted for the mean and standard deviation of the values within a window because these statistics effectively capture the basic characteristics of the signal. The mean, representing central tendency, provides insights into average signal amplitude, while the standard deviation quantifies the dispersion of values, indicating signal variability. This combination offers a concise summary of EEG dynamics while being computationally efficient and suitable for reducing dimensionality. This approach strikes a balance between simplicity, informative representation, and classification efficacy in EEG signal analysis. The second approach of this study involved dividing the EEG frequency spectrum into five distinct subbands based on brain rhythms: delta ($0.5-4$ Hz), theta ($4-8$Hz), alpha ($8-14$ Hz), beta ($14-30$ Hz), and gamma (30-45 Hz) [33]. We are utilizing the complete range of these bands, which are commonly associated with various cognitive processes, to capture a broader spectrum of neural activity. From these subbands, the power spectral density (PSD) values were extracted. Although traditional methods frequently use joint time-frequency analysis techniques, such as Continuous Wavelet Transform (CWT) and Stockwell transforms, we prioritize PSD analysis to effectively capture neural activity across the identified frequency bands. Previous research has consistently shown the significance of these specific frequency bands in characterizing different mental states [34] [35].

ECG and EDA. Regarding ECG, a typical heartbeat (cardiac cycle) lasts around 0.8 seconds [36]. Using a 1-second window size would restrict the analysis to capture only one complete cardiac cycle. Therefore, 10-second windows were used. Concerning EDA signals, window size selection is more flexible. While previous research on EDA has utilized window sizes ranging from 10 to 300 seconds [37] [38], we opted for a 10-second window in this analysis. This choice aligns with the window size used for ECG signals and allows for the integration of these two different types of physiological data.

3.4 Labeling and Class Balancing

Initially, a binary labeling scheme was employed to classify each sample as either "stress" or "relax." However, due to instances where the subject's state was ambiguous, a

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

third class called ”neutral” was introduced. This new class accounted for time periods where the subject’s emotional state was neither distinctly stressed nor relaxed, providing a more accurate representation of their responses during the experiment.

Labeling has been conducted by visual inspection. To perform the visual labeling, references were derived from the recorded ECG and EDA data obtained with the Biopac device. Both ECG and EDA signals are known to reflect changes in the autonomic nervous system, which is closely linked to emotional responses. The ECG signal stabilizes during relaxed states, while the EDA signal decreases, indicating a reduction in the subject’s sweating. Conversely, in stressful situations, the EDA signal tends to increase due to heightened perspiration. Fig. 7 shows a fragment of the captured ECG and EDA signals. As can be seen, the ECG signal stabilizes when the subject enters a relaxed state. At the same time, the EDA signal decreases in that state, indicating a decrease in the subject’s sweating, while in the state of stress, it tends to increase [39]. By analyzing these physiological indicators, relevant time periods during the experiment were identified for labeling as ”stress”, ”relax” or ”neutral.”

The labels assigned to the ECG and EDA data based on the physiological responses were also extended to the simultaneous EEG data. Since all three signals were acquired simultaneously, their temporal alignment allowed for the association of the emotional states determined from ECG and EDA signals with the corresponding EEG data samples. This integration ensured that the emotional labels derived from the physiological data were consistently applied to the EEG data, thus enabling a comprehensive classification of the EEG signals into the three categories.

The information from questionnaires and interviews gathered from the subjects helped to refine the labeling process and ensure that the emotional states assigned to each data sample aligned more accurately with the participants’ actual experiences during the stress induction tests. The questionnaires consisted of a simplified version of the SAM [40], focusing solely on the questions. By combining the objective physiological data from ECG and EDA signals with subjective feedback from post-experiment interviews and the streamlined SAM questionnaire, a more robust and nuanced visual labeling approach was established.

Fig. 7: Labeling with ECG and EDA signals

After labeling, the number of samples in each class was not evenly distributed. Imbalanced class distributions can lead to biased model training, where the model might favor the majority class, resulting in suboptimal performance in detecting minority classes. To address this issue, a class balancing technique, specifically random oversampling, was utilized. Random oversampling involves duplicating instances from the minority classes, thereby increasing their representation in the dataset and achieving a more balanced distribution across all classes [41]. Although random oversampling might cause the model to become overly reliant on the duplicated minority class instances, we have preferred this technique to other ones that generate synthetic samples as SMOTE [42] because the latter is based on interpolation between existing instances. In some cases, this process may introduce noise or outliers, impacting the overall quality of the dataset and potentially affecting model performance. Moreover, methods like SMOTE assume local linearity between instances, while in EEG data, where complex nonlinear relationships may exist, this assumption might not always hold, potentially leading to the generation of unrealistic synthetic samples.

3.5 Classification

We considered LightGBM, a gradient-boosting model with a tree-based learning algorithm, as the first model for our classification task due to its promising results in the literature [43] [44]. Additionally, we also explored other algorithms, such as a state-of-the-art 16-layer CNN [45], KNN (with Grid Search for Hyperparameter Tuning), and a Radial Basis Function Support Vector Machine (RBF-SVM). The chosen algorithms span a spectrum of complexity. LightGBM and CNN are capable of capturing complex relationships, KNN is simple and non-parametric, while SVM creates optimal hyperplanes, accommodating various data complexities. In particular, RBF kernels have shown superior performance in various scenarios [46]. These diverse models were trained on our datasets to comprehensively evaluate their effectiveness in stress detection.

3.6 Experimental Setup

The main aim of our work is to analyze the contribution of EEG for stress detection. To that end, we defined five experimental scenarios.

1) **EEG Data for Stress Detection**. First, we define several combinations of features to determine the best set of features that provides the highest performance for stress detection over EEG data according to the machine learning methods proposed.
2) **Sensitivity Analysis of the Window Size on EEG**. As feature extraction depends on the window size, a sensitive analysis is conducted regarding this parameter of the methodology.
3) **EEG Intrasubject Analysis**. An intrasubject classification was conducted to assess the feasibility of creating personalized models without sacrificing accuracy. This analysis aimed to identify any conflicts in model classification caused by individual subject samples.
4) **Potential of EEG regarding ECG and EDA**. We analyze the potential of EEG data regarding ECG

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

and EDA for stress detection, by combining different configurations of the biosignals. This experiment would validate the main contribution of this work. Our hypothesis is that EEG provides richer information regarding ECG and EDA.

5) **EEG Channel Importance**. In pursuit of simpler EEG-based stress detection methods, we analyze which EEG channels offer the greatest contribution to identifying stress.

All models were trained using 80% of the available data, while the remaining 20% was reserved for testing purposes. To validate the results, a 5-fold cross-validation has been applied to ensure that the models were rigorously evaluated to obtain robust and reliable performance estimates. For the intersubject validation, we employed separate subjects for training and testing to ensure that our model's performance is robust and generalizable across different individuals. Regarding intrasubject validation, we ensured that the windows do not overlap, which inherently means that there is no shared information between the training and testing sets. This approach minimizes the risk of data leakage and helps maintain the integrity of the validation process. The classification performance was evaluated using accuracy, precision, and F1 score. Other particular details of every experiment configuration are provided below.

3.6.1 EEG Data for Stress Detection: Configuration Set Up

Three feature sets were extracted from the preprocessed EEG data: "Mean & std", "Frequency bands", and "Mean & std + bands". Lastly, the "Mean & std + bands" feature set combined both statistical and frequency domain features, aiming to capture a broader range of information. The window size used is 1 second according to the methodology explained in Section 3.3.

Regarding machine learning, first, a three-class problem has been considered (stress, neutral, relax), and next a one-vs-all classification approach was employed to distinguish stress vs. all other states and relaxation vs. all other states. The binarization of the classification task with the one-vs-all strategy is used to provide more accurate results.

3.6.2 Sensitivity Analysis of the Window Size on EEG: Configuration Set Up

Feature selection results are subject to the window size. Therefore, different window sizes (1s, 10s, and 20s) were considered. The experiment was performed using the LightGBM machine learning algorithm as the sole model for the classification task, according to our proposal (see Section 3.5).

3.6.3 EEG Intrasubject Analysis: Configuration Set Up

Personalized or Precision Medicine is a concept that relies on building models based on individual, rather than multi-subject data. The key idea is to base medical decisions on individual patient characteristics rather than population averages to achieve the best possible outcome through personalized treatments [47]. For this reason, an intrasubject classification was performed to analyze whether it is feasible to create personalized models without losing accuracy and to check whether any subject's samples cause conflicts in the classification of the models. For each subject, the experiment was performed according to our proposal (see Section 3.5), using the "Mean & std" feature set and the LightGBM algorithm.

3.6.4 Potential of EEG regarding ECG and EDA: Configuration Set Up

To ensure a uniform temporal representation of the physiological signals, a 10-second window size was chosen for feature extraction, as explained in Section 3.3. This decision was further influenced by the favorable results obtained from window size analysis with EEG data in the previous experiments.

The features extracted from the EEG data were the mean and standard deviation values computed from the 19 EEG channels. Additionally, the mean and standard deviation of the ECG and EDA signals were also extracted as features. These features were chosen due to their proven effectiveness in representing the underlying physiological patterns relevant to stress detection in the previous experiments. The machine learning approach used is LightGBM.

3.6.5 EEG Channel Importance: Configuration Set Up

This analysis aimed to study the EEG channels (see Fig. 2) that provide the most information for predicting stress states, quantifying the feature contributions as percentages of all channels. The ultimate goal was to identify the existence of specific areas that are more relevant for stress detection, potentially allowing the replacement of the electrode cap with a less intrusive portable EEG device (i.e. wearable).

The quantification of the feature contributions has been conducted with the feature importance method provided by the LightGBM method. Feature importance in LightGBM is calculated using two main approaches: split importance and gain importance. Split importance counts how frequently a feature is used to split the data across all decision trees in the model, while gain importance measures the average improvement in the model's performance achieved by splitting on a feature. After calculating these importance scores, they are normalized to bring them to a common scale. The combined feature importance score is then obtained by averaging the normalized scores from both methods. This combined score provides insights into which features, in this case, channels/electrodes, are most influential in predicting the target variable.

The features used in this experiment were "Mean & std" feature set and a window size of 10 sec, consistent with the previous experimental results.

4 RESULTS

This section presents the findings obtained from the analysis of the results obtained with the experimentation conducted in this study.

4.1 Results on EEG Data for Stress Detection

Table 2 compares classification results using different feature sets and machine learning models based on EEG signals. The accuracy of the models is provided for each feature-model combination.

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

Feature set	LightGBM			CNN			KNN			SVM		
	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1
Mean & std	**86.24**	**84.12**	**84.15**	51.57	50.23	50.25	82.00	81.66	81.73	55.98	55.06	55.11
Frequency bands	63.18	67.56	67.63	32.70	33.39	33.42	70.00	72.14	72.21	35.06	37.56	37.62
Mean & std + bands	84.34	83.82	83.86	52.02	51.67	51.75	81.00	80.53	80.61	52.36	50.98	51.06

TABLE 2: Classification results: Comparison of features and models

The ”Mean & std” feature set achieved the highest accuracy (86.24%) with LightGBM, while ”Frequency bands” had the lowest accuracies for all models. The CNN and SVM machine learning approaches performed worse compared to LightGBM and KNN.

The ”Mean & std” feature set performed consistently better across different models, demonstrating its effectiveness for classification. The ”Frequency bands” feature set showed lower accuracies, indicating that it might not capture enough discriminative information on its own. The combination of ”Mean & std + bands” yielded competitive results, although slightly lower than ”Mean & std” for LightGBM, KNN and SVM. Overall, ”Mean & std” appears to be the most reliable and informative feature set.

Comparing the four models, LightGBM is validated as the most effective model for the classification task (relax, neutral, stress), achieving the highest accuracies across two different feature sets. KNN performed reasonably well, while CNN and SVM had the least success in accurately classifying the EEG signals.

Table 3 shows classification results when considering the binarization problem, in which the one-vs-all strategy has been applied. Therefore, two classification tasks have been defined: distinguishing stress vs. all other states and relaxation vs. all other states. LightGBM performed best in both tasks, achieving the highest accuracy for relaxation vs. all (93.89%) and competitive accuracy for stress vs. all (86.77%). On the other hand, KNN also performed well but fell short of LightGBM's accuracy, while CNN and SVM had lower accuracies in both tasks.

The results consistently demonstrate that LightGBM is the most effective machine learning model across different scenarios. Not only did it excel in the classification task using the best feature set (”Mean & std”), achieving the highest accuracy, but it also outperformed other models when following a one-vs-all strategy. This reinforces the superiority of LightGBM in accurately classifying EEG signals for stress detection in this study.

The computational cost of feature extraction is also crucial for understanding a classification process's efficiency, particularly in real-time applications or those dealing with large datasets. Table 4 presents the average time taken for feature extraction for each feature set.

As the results show, ”Mean & std” is significantly faster (588 ms) compared to ”Frequency bands” (1,123 ms). This substantial difference underlines that calculating mean values is computationally less demanding than the broader feature extraction process. These findings emphasize the importance of considering the computational cost of feature extraction during feature selection.

4.2 Results on Sensitivity Analysis of the Window Size on EEG

Table 5 displays classification results for different feature sets and window sizes. ”Mean & std” consistently outperforms other feature sets in accuracy, precision, and F1 scores. ”Frequency bands” has the lowest performance. Combining ”Mean & std” with ”Frequency bands” shows some improvement but still falls short of ”Mean & std” alone.

For the ”Mean & std” feature set, the performance drops slightly when transitioning from a 1s window size to larger window sizes of 10s and 20s. This indicates that a smaller window size provides more fine-grained information, leading to better classification results for this particular feature set. In contrast, for the ”Frequency bands” feature set, there is a consistent improvement in accuracy, precision, and F1 scores with increasing window size. The ”Mean & std + bands” feature set also exhibits a similar pattern as the ”Mean & std” feature set, with diminishing performance as window size increases.

The contrasting behavior of the ”Frequency bands” and ”Mean & std” feature sets with respect to window size can be attributed to the types of information they capture. Larger window sizes benefit the ”Frequency bands” set, as they allow the model to capture broader patterns in frequency distribution. Conversely, the ”Mean & std” set may be adversely affected by larger window sizes, losing sensitivity to rapid variations and key patterns.

4.3 Results on EEG Intrasubject Analysis

Figure 8 presents the intrasubject variation in accuracy scores obtained across different subjects. The dashed red line denotes the average accuracy of the LightGBM model across all subjects, which stands at 76.92%.

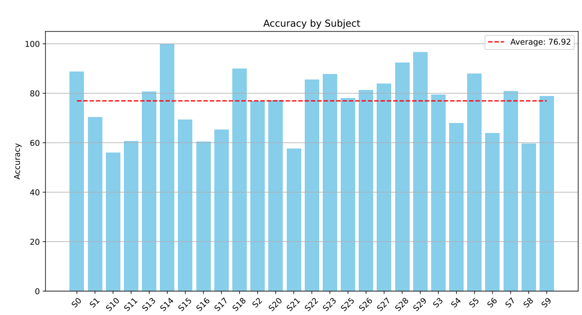

Fig. 8: Classification results: Intrasubject

Comparing the intrasubject to the intersubject results, where an accuracy of 86.24% was achieved, reveals notable differences. The intrasubject analysis highlights considerable variability in model performance across individual

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

	LightGBM			CNN			KNN			SVM		
	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1
Stress vs the rest	86.77	86.54	86.67	69.03	67.98	68.05	77.00	78.22	78.27	63.41	61.09	61.13
Relax vs the rest	**93.89**	**93.94**	**93.91**	51.08	50.12	50.20	77.00	77.25	77.25	63.41	61.76	61.76

TABLE 3: Classification results: One vs All

Feature set	Time (ms)
Mean & std	**588**
Frequency bands	1,123

TABLE 4: Average computational times

Feature set	Window size	Acc	Prec	F1
	1s	**86.24**	**84.12**	**84.15**
Mean & std	10s	85.13	86.51	86.53
	20s	83.55	88.90	88.92
	1s	63.18	67.56	67.63
Frequency bands	10s	63.40	69.37	69.47
	20s	65.44	69.59	69.64
	1s	84.34	83.82	83.86
Mean & std + bands	10s	84.20	83.78	83.80
	20s	83.25	81.23	81.22

TABLE 5: Classification results: Window size

Biosignals	Acc	Prec	F1
EEG	85.13	86.51	86.53
ECG + EDA	75.88	75.46	75.04
EEG + ECG + EDA	**91.44**	**91.54**	**91.42**

TABLE 6: Classification results: Comparison between EEG and ECG + EDA data.

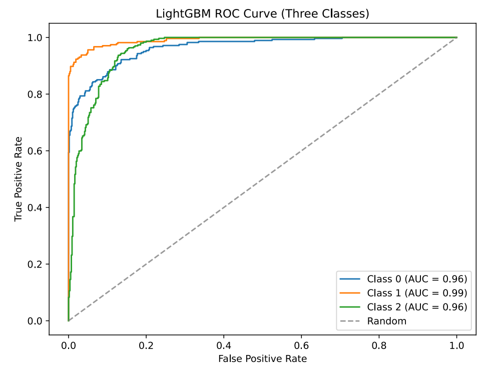

Fig. 9: ROC Curve: EEG (0: relax, 1: neutral, 2: stress)

subjects, with accuracy scores ranging from near-perfect (e.g., subjects S14 and S18) to relatively lower scores (e.g., subjects S10 and S21).

The observed intrasubject variability may stem from individual differences in physiological responses or inherent noise within the EEG signals. While the intersubject analysis demonstrates the strong performance of the ”Mean & std” feature set, the intrasubject analysis underscores the importance of considering individual variability in model evaluation. Despite the variability, the LightGBM model maintains a relatively high average accuracy across most of the subjects.

4.4 Results on the Potential of EEG regarding ECG and EDA

Table 6 presents the classification results comparing different combinations of signals: EEG alone, ECG and EDA, and all three biosignals.

The results show that EEG data improves up to 10 points the results of ECG and EDA together (12.19%). This could be the reason why there are most recent research works focusing on EEG. However, it is important to highlight that EEG, ECG, and EDA data combined leads to significantly improved classification performance compared to using individual biosignals alone: the combined feature set provides six points more accuracy (6.31%).

In addition to the tabulated results, the analysis of Receiver Operating Characteristic (ROC) curves further elucidates the discriminatory power of the classifiers. Specifically, when considering individual modalities, the EEG-based classifier (Fig. 9) demonstrates strong performance, with area under the curve (AUC) values of 0.96 for Class 0 (relax), 0.99 for Class 1 (neutral), and 0.96 for Class 2 (stress). Similarly, the ECG + EDA-based classifier (Fig. 10) exhibits respectable AUC values of 0.91, 0.95, and 0.90 for Classes 0, 1, and 2, respectively. Remarkably, combining EEG with ECG and EDA (Fig. 11) yields further improvements, with AUC values of 0.98, 1.00, and 0.98 for Classes 0, 1, and 2, respectively. These findings underscore the synergistic benefits of integrating multiple modalities, as evidenced by the enhanced discriminatory power observed in the combined classifier. However, the small differences should be discussed in depth, regarding the implications of the use of a single or multiple devices.

4.5 Results on EEG Channel Importance

As explained in Section 3.6.5, this analysis focused on determining the importance of various EEG channels for predicting stress, particularly by examining the mean and standard deviation. Fig. 12 presents the feature importance values obtained using LightGBM for all the EEG electrodes. ”T8” exhibits the highest importance with a value of 0.065643, followed closely by ”Fp1”, ”Fz” and ”C3” with importance values of 0.064449, 0.063147 and 0.063115, respectively. Other notable electrodes include ”Pz”, ”F8”, ”F3” and ”P4” (0.059876 to 0.057654). ”P7”, ”O2”, ”P8”, ”T7” and ”F4” have lower but meaningful contributions (0.055432 to 0.051234). Lastly, ”O1,” ”F7,” ”P3,” ”C4,” and ”Fp2” are the least important (0.050123 to 0.043510).

This article has been accepted for publication in IEEE Transactions on Affective Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2024.3503995

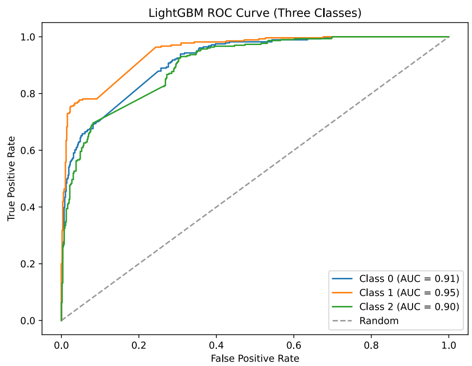

Fig. 10: ROC Curve: ECG + EDA (0: relax, 1: neutral, 2: stress)

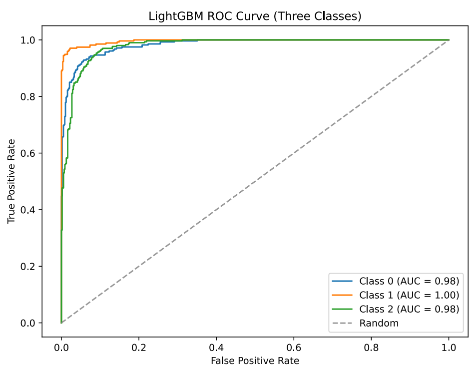

Fig. 11: ROC Curve: EEG + ECG + EDA (0: relax, 1: neutral, 2: stress)

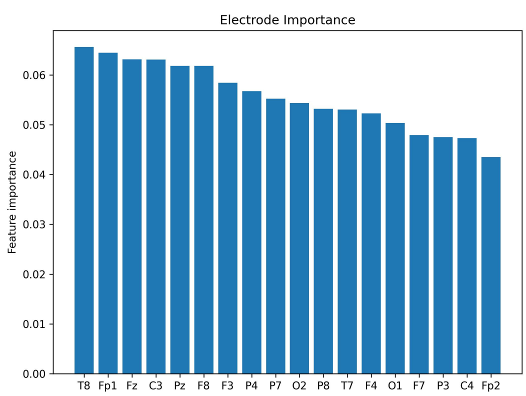

Fig. 12: Electrode Importance

5 Discussion

The results show that EEG is a good source of data for stress detection; this is something that previous works already support. ECG and EDA are behind of the EEG potential, performing up to a 12.19% worse. Although the combination of all of the sensors enhances stress detection (7.41% concerning EEG alone), the results should be understood in the context of where the data is gathered, and the critical condition of the patient. In the case of considering wearable devices capable of collecting diverse physiological signals for stress detection applications, could be a difficult endeavor. ECG and EDA could be available in a single product (e.g. wristband), while EEG could be in a different one (for example the brain–sensing Muse headband used in [16]). Choosing one or both could respond to the severity condition of the subject. Therefore, this work contributes to understanding the information provided by each sensor, and to choosing the appropriate combination according to the monitoring conditions.

In terms of the research methodology, this work uses an inter-subject and intra-subject design. In the inter-subject design, we gather data from multiple participants experiencing stress through controlled activities. Consequently, the findings are generalizable to a broader population. Regarding the intra-subject design, data labeling has been performed according to the ECG and EDA signals. This is of particular importance since most previous works use EEG data alone, and the stress state is determined according to the protocol procedure, and not to the individual reaction. The chosen duration might not induce uniformly strong stress responses in all participants. For instance, the 7-minute puzzle could allow participants to adapt to the challenge over time, potentially reducing stress levels as they progress. We might need to consider incorporating puzzles with increasing difficulty or monitor stress levels throughout the activity to capture the initial stress response more effectively. The 3-minute mental math calculation, on the other hand, could introduce fatigue towards the end, impacting the reliability of the collected physiological data.

Concerning the features to be considered to represent the biosignals for stress detection, this work discloses the performance of simple features, as the ”Mean & std”. Using statistical measures that are easy to compute and a fast model like LightGBM offers several advantages for stress detection. LightGBM is known for its efficiency in handling large datasets and rapid training times, making it a relatively simple and fast approach to classifying EEG data for stress detection. Its speed and low computing cost allow for quick real-time analysis, making it suitable for applications requiring timely stress monitoring.

Analyzing the channel importance, the results show that despite the minor disparities in the percentage values, some differences suggest that certain electrodes offer more valuable information than others for stress detection. This finding holds particular significance in fields like Brain-Computer Interface (BCI) research. There are wearable BCI devices in the market with low visual impact, that could be applied for stress detection. The ability to detect stress through a focused brain zone could be especially relevant in specific areas such as surgery or other clinical applications. For example, during surgical procedures, the detection of stress in the operating room personnel, such as surgeons or nurses, could be valuable in ensuring optimal performance and patient safety.

5.1 Limitations

This study has provided valuable insights into stress detection using EEG data, but it is essential to recognize some of its limitations. Firstly, in the intrasubject analysis, the current model employs a one-size-fits-all strategy, uniformly treating everyone's data. While this approach provides a general overview of performance across a diverse population, it may not fully capture individual variations within the data. To address this limitation, alternative strategies focusing on personalized modeling techniques can be explored. Secondly, conducting the experiment with a more diverse group of subjects would bolster the validity and applicability of the findings. Moreover, the study might have been conducted under specific conditions or in a controlled environment that may not fully represent real-life scenarios. Replicating the experiment under various settings and contexts would offer a more comprehensive understanding of stress detection in different situations.

6 Conclusion

The field of stress detection using physiological signals has gathered significant interest in recent years due to its potential for enhancing mental health monitoring and well-being [48]. In this work, we focus on the study of the role of EEG data for stress detection. A data collection protocol has been defined in order to induce stress in the participants, and the biosignals have been gathered with EEG, EDA, and ECG. Different feature extraction methods and machine learning models have been tested to accurately identify stress states.

The results of our experiments emphasize the significance of selecting appropriate feature sets, machine learning models, and window sizes for stress detection based on EEG data. LightGBM consistently proved to be the most effective method, achieving the highest accuracies across different scenarios. EEG data outperforms ECG and EDA, but all the physiological signals together achieve the best outcome. Nevertheless, the integration of multiple physiological signals depends on the context in which stress would be measured.

Future research includes exploring contextual information on stress detection. The investigation of stress detection in more ecologically valid and real-world scenarios can also be an important direction for future studies. Furthermore, the application of attention mechanisms might further enhance the classification performance [49].

Acknowledgments

The authors would like to express their gratitude to all the participants who generously contributed their time and cooperation to this study. This study was conducted with the support of the Generalitat de Catalunya, via the Consolidated Research group 2021 SGR 01125.

References

[1] World Health Organization. (2023, February 21). Questions and answers on stress. Retrieved from https://www.who.int/news-room/questions-and-answers/item/stress (accessed on March 26, 2024).

[2] D. Lecic-Tosevski, O. Vukovic and J. Stepanovic, "Stress and personality," *Psychiatrike = Psychiatriki*, 22(4), 290–297, 2011.

[3] P. E. Flores-Kanter, L. Moretti, and L. A. Medrano, "A narrative review of emotion regulation process in stress and recovery phases," *Heliyon*, vol. 7, no. 6, Jun. 2021, p. e07218, doi: 10.1016/j.heliyon.2021.e07218.

[4] A.D. Crosswell, K.G. Lockwood, "Best practices for stress measurement: How to measure psychological stress in health research," *Health Psychol Open*, 2020, doi: 10.1177/2055102920933072.

[5] R. Martínez, E. Irigoyen, A. Arruti, J.I. Martin and J. Muguerza, "A Real-Time Stress Classification System based on Arousal Analysis of the Nervous System by an F-State Machine," *Computer Methods and Programs in Biomedicine*, vol. 148, 2017.

[6] R. Nawaz, K. H. Cheah, H. Nisar and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 910–926, 2020.

[7] F. Agrafioti, D. Hatzinakos and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 102–115, First Quarter 2012.

[8] R. Martinez, A. Salazar-Ramirez, A. Arruti, E. Irigoyen, J. I. Martin and J. Muguerza, "A Self-Paced Relaxation Response Detection System Based on Galvanic Skin Response Analysis," in *IEEE Access*, vol. 7, pp. 43730-43741, 2019.

[9] Y. Shan, S. Li and T. Chen, "Respiratory signal and human stress: non-contact detection of stress with a low-cost depth sensing camera," *International Journal of Machine Learning and Cybernetics*, 2020, doi: 10.1007/s13042-020-01074-x.

[10] S. A. Patil and A. N. Paithane, "Comprehensive Review of EEG-Based Algorithms for Mental Stress Analysis," *IEEE International Conference on Computational Intelligence and Ubiquitous Computing for Biomedical Applications (CIBEC)*, 2023, pp. 1-7, doi: 10.1109/ICCUBEA58933.2023.10392207.

[11] M.B. Bin Heyat, F. Akhtar, S.J. Abbas, M. Al-Sarem, A. Alqarafi, A. Stalin, R. Abbasi, A.Y. Muaad, D. Lai and K. Wu, "Wearable Flexible Electronics Based Cardiac Electrode for Researcher Mental Stress Detection System Using Machine Learning Models on Single Lead Electrocardiogram Signal," *Biosensors*, 2022, 12, 427. https://doi.org/10.3390/bios12060427

[12] E. Hosseini et al., "A Low Cost EDA-based Stress Detection Using Machine Learning," *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Las Vegas, NV, USA, 2022, pp. 2619-2623, doi: 10.1109/BIBM55620.2022.9995093.

[13] A. Asif, M. Majid and S. M. Anwar, "Human stress classification using EEG signals in response to music tracks," *Comput. Biol. Med.*, vol. 107, pp. 182-196, 2019.

[14] C. G. Ávila, F. S. Bott, L. Tiemann, V. D. Hohn, E. S. May, M. M. Nickel, P. T. Zebhauser, J. Gross, and M. Ploner, "DISCOVER-EEG: an open, fully automated EEG pipeline for biomarker discovery in clinical neuroscience," *Scientific Data*, vol. 10, no. 1, Sep. 2023, p. 613, doi: 10.1038/s41597-023-02525-0.

[15] B. Gosala, P. D. Kapgate, P. Jain, R. N. Chaurasia, and M. Gupta, "Wavelet transforms for feature engineering in EEG data processing: An application on Schizophrenia," *Biomedical Signal Processing and Control*, vol. 85, Aug. 2023, p. 104811, doi: 10.1016/J.BSPC.2023.104811.

[16] N. Armanfard, M. Komeili, J. P. Reilly, and L. Pino, "Vigilance lapse identification using sparse EEG electrode arrays," *Canadian Conference on Electrical and Computer Engineering (CCECE)*, Oct. 2016, doi: 10.1109/CCECE.2016.7726846.

[17] S. Samarpita, R. Satpathy, B. Mishra and R. Mishra, "Differentiating Mental Stress Levels: Analysing Machine Learning Algorithms Comparatively For EEG-Based Mental Stress Classification Using MNE-Python," *Journal of Advanced Zoology*, 2023, doi: 10.17762/jaz.v44iS5.2013.

[18] S. Baliga, N. Ali, A. LS and V. P, "Emotion Recognition and Stress Reduction Based on Electroencephalograph (EEG) Signals validated by Machine Learning Algorithms," *2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, Tumakuru, India, 2023, pp. 1-6, doi: 10.1109/ICSSES58299.2023.10199291.

[19] A.J. Marthinsen, I.T. Galtung1, A. Cheema1, C.M. Sletten1, I.M. Andreassen1, Øystein Sletta1, A. Soler1 and M. Molinas, "Psychological stress detection with optimally selected EEG channel using Machine Learning techniques," *CEUR Workshop Proceedings*, 2023, 53-68.

[20] O. AlShorman, M. Masadeh, Md B.B. Heyat, F. Akhtar, H. Almahasneh, G. Md Ashraf and A. Alexiou, "Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection," *J. Integr. Neurosci.*, 2022, 21(1), 20.

[21] A. Nirabi, F. A. Rahman, M. H. Habaebi, K. A. Sidek and S. Yusoff, "Machine Learning-Based Stress Level Detection from EEG Signals," *2021 IEEE 7th International Conference on Smart Instrumen-*

tation, Measurement and Applications (ICSIMA), Bandung, Indonesia, 2021, pp. 53-58, doi: 10.1109/ICSIMA50015.2021.9526333.

[22] A. Hemakom, D. Atiwiwat and P. Israsena, ”ECG and EEG based detection and multilevel classification of stress using machine learning for specified genders: A preliminary study,” *PLoS One*, 2023 Sep 1;18(9):e0291070. doi: 10.1371/journal.pone.0291070.

[23] P. Zontone, A. Affanni, R. Bernardini, A. Piras and R. Rinaldo, ”Stress Detection Through Electrodermal Activity (EDA) and Electrocardiogram (ECG) Analysis in Car Drivers,” *2019 27th European Signal Processing Conference (EUSIPCO)*, A Coruña, Spain, 2019, pp. 1-5, doi: 10.23919/EUSIPCO.2019.8902631.

[24] A. Affani, ”Wireless Sensors System for Stress Detection by Means of ECG and EDA Acquisition,” *Sensors*, vol. 20, no. 7, 2020, doi: 10.3390/s20072026.

[25] T. Umematsu, A. Sano, S. Taylor and R. W. Picard, ”Improving Students’ Daily Life Stress Forecasting using LSTM Neural Networks,” *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Chicago, IL, USA, 2019, pp. 1-4, doi: 10.1109/BHI.2019.8834624.

[26] P. Bobade and M. Vani, ”Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data,” *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2020, pp. 51-57.

[27] J. Shin, J. Maeng and D.-H. Kim, ”Inner emotion recognition using multi bio-signals,” in *Proc. IEEE Int. Conf. Consum. Electron.-Asia*, 2018, pp. 206–212.

[28] M. Norden, AG Hofmann, M. Meier, F. Balzer, OT Wolf, E. Böttinger and H. Drimalla, ”Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study,” *J Med Internet Res*, 2022, doi: 10.2196/32280.

[29] M.A. Almazrouei, R.M. Morgan and I.E. Dror, ”A method to induce stress in human subjects in online research environments,” *Behav Res 55*, 2023, doi: 10.3758/s13428-022-01915-3.

[30] M. Klug and K. Gramann, ”Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments,” *The European journal of neuroscience*, vol. 54, 2021.

[31] L. Sun, Y. Liu and P. J. Beadle, ”Independent component analysis of EEG signals,” *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology*, pp. 219-222, 2005.

[32] F. Perrin, J. Pernier, O. Bertrand and J. F. Echallier, ”Spherical splines for scalp potential and current density mapping,” *Electroencephalography and Clinical Neurophysiology*, vol. 72, Issue 2, 1989.

[33] C. Beste, A. Münchau and C. Frings, ”Towards a systematization of brain oscillatory activity in actions,” *Communications Biology*, 6(1), pp. 1–11, 2023, doi: 10.1038/s42003-023-04531-9.

[34] E. Perez-Valero, M. Lopez-Gordo and M. Vaquero-Blasco, ”EEG-based multi-level stress classification with and without smoothing filter,” *Biomedical Signal Processing and Control*, vol. 69, 2021, doi: 10.1016/j.bspc.2021.102881.

[35] M. A. Hafeez and S. Shakil, ”EEG-based stress identification and classification using deep learning,” *Multimedia Tools and Applications*, vol. 83, pp. 1-17, 2023, doi: 10.1007/s11042-023-17111-0.

[36] H. Bian, Y. Bian, J. Li, S. Xu, X. Shao, J. Li and B. Jiang, ”A simple hand mnemonic for teaching the cardiac cycle,” *Advances in Physiology Education*, vol. 44, no. 1, 2019.

[37] A. de Santos Sierra, C. S. Ávila, J. G. Casanova, and G. B. del Pozo, ”A stress-detection system based on physiological signals and fuzzy logic,” *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4857–4865, Oct. 2011, doi: 10.1109/TIE.2010.2103538.

[38] J. A. Healey and R. W. Picard, ”Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[39] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón and K. H. Chon, ”Electrodermal Activity Is Sensitive to Cognitive Stress under Water,” *Front Physiol.*, 2018, Jan 17; 8:1128, doi: 10.3389/fphys.2017.01128.

[40] TM. Bynion, M.T. Feldner, ”Self-Assessment Manikin,” *Encyclopedia of Personality and Individual Differences*, Springer, 2017, doi: 10.1007/978-3-319-28099-8_77-1.

[41] H. Ali, M. Salleh, K. Hussain, A. Ullah, A. Ahmad and R. Naseem, ”A review on data preprocessing methods for class imbalance problem,” *International Journal of Engineering & Technology*, 390-397, 2019, doi: 10.14419/ijet.v8i3.29508.

[42] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, ”SMOTE: synthetic minority over-sampling technique,” *J. Artif. Int. Res.*, vol. 16, 2002, 321–357.

[43] H. Zeng, C. Yang, H. Zhang, W. Zhenhua, J. Zhang, G. Dai, F. Babiloni and W. Kong, ”A LightGBM-Based EEG Analysis Method for Driver Mental States Classification,” *Computational Intelligence and Neuroscience*, 2019. 1-11. 10.1155/2019/3761203.

[44] S. Aggarwal, L. Aggarwal, M. S. Rihal and S. Aggarwal, ”EEG Based Participant Independent Emotion Classification using Gradient Boosting Machines,” *2018 IEEE 8th International Advance Computing Conference (IACC)*, Greater Noida, India, 2018, pp. 266-271, doi: 10.1109/IADCC.2018.8692106.

[45] D. Acharya et al., ”Multi-class Emotion Classification Using EEG Signals,” *Communications in Computer and Information Science*, vol. 1367. Springer, IACC 2020.

[46] Q. Que and M. Belkin, ”Back to the Future: Radial Basis Function Network Revisited,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 1856-1867, 1 Aug. 2020, doi: 10.1109/TPAMI.2019.2906594.

[47] H. Fröhlich et al., ”From hype to reality: data science enabling personalized medicine,” *BMC Med.*, 2018, doi: 10.1186/s12916-018-1122-7.

[48] T. Iqbal, A.J. Simpkin, D. Roshan, N. Glynn, J. Killilea, J. Walsh, G. Molloy, S. Ganly, H. Ryman, E. Coen et al. ”Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset,” *Sensors*, 2022, doi: 10.3390/s22218135.

[49] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao and Y. Zhang, ”EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network,” *Knowledge-Based Systems*, vol. 265, 2023.

Jonah Fernandez received the Bachelor’s degree in Automation and Industrial Electronics Engineering and the MSc degree in Data Science from the University of Girona (UdG), Spain, in 2020 and 2022, respectively. He is currently working toward the PhD degree with the University of Girona. His research interests include neuroengineering, neurological diseases detection and machine learning.

Raquel Martinez received the M.Sc. degree in electronic and automation engineering in 2003, and the Ph.D. in System Engineering and Automation in 2016, from the University of the Basque Country. She is an Assistant Professor with the Department of System Engineering and Automation, University of the Basque Country. She leads the Algorithms Data mining and Parallelism (ALDAPA) research group. Her research interests include physiological computing, experimental setup, among others.

Bianca Innocenti is an assistance professor at the Department of Electrical, Electronics and Automation Engineering of the University of Girona. Her research area is artificial intelligence applied to robotics and healthcare, emphasizing on data acquisition.

Beatriz López is a professor at the Department of Electrical, Electronics and Automation Engineering of the University of Girona, director of the research lab on AI applied to Medicine and Healthcare in the eXiT research group. She has published 3 books and more than 200 articles in magazines and conferences. She was president of the Catalan Association of AI (member of EurAI.org) during 2014-2016, and vice-president 2010-2014.