Universitat de Girona
Escola Politècnica Superior

Treball Final de Màster

Estudi: Màster en Ciència de Dades

Títol: Human stress detection using EEG signals

Document: Memòria

Alumne: Jonah Fernandez Gonzalez

Tutor: Dra. Beatriz López Ibáñez
Departament: Enginyeria Elèctrica, Electrònica i Automàtica
Àrea: Enginyeria de Sistemes i Automàtica

Tutor (extern): Dra. Raquel Martínez Rodríguez

Convocatòria (mes/any): Juny 2022

Universitat de Girona
Escola Politècnica Superior

MASTER'S THESIS

Human stress detection using EEG signals

Author:
Jonah FERNANDEZ GONZALEZ

June 2022

Master in Data Science

Tutors:
Beatriz LÓPEZ IBÁÑEZ
Raquel MARTÍNEZ RODRÍGUEZ

Abstract

This Master Thesis aimed to detect stress and relaxation states from EEG data acquired with the Neuroelectrics Enobio device and contrast them with other physiological signals, in this case the electrocardiogram (ECG) and galvanic skin response (GSR) acquired with the Biopac MP36 system. In the literature, there are several investigations conducted in this field using EEG or ECG/GSR signals, but there are much fewer studies assessing the benefits of combining EEG signals with other physiological signals for mental state classification.

At the same time, the performance of different Machine Learning and Deep Learning techniques were investigated to corroborate which one is the most suitable to achieve the proposed goals. The models used were LightGBM, a 16-layer CNN, a KNN with Grid Search for Hyperparameter Tuning and a non-linear SVM.

The data used in this work have been obtained from students at the University of Girona. To acquire data from the subjects, an inductive stress experiment was conducted in a controlled work environment. Each subject underwent several tests with relaxation videos interspersed between each test. The data were labeled into three classes (*stress*, *relax* or *neutral*) based on the values of the ECG and GSR signals and contrasted with the post-experiment survey conducted on each subject. To balance the ratio of the three classes, random oversampling was applied to the data set.

Each model was trained in several phases with different scenarios: calculating only the means and standard deviations of each sample window, extracting the five frequency bands of the EEG signals (delta, theta, alpha, beta and gamma), segmenting the data in windows (with and without overlapping), applying a binary problem of stress/relax vs the rest, using only the EEG or ECG/GSR signals and the combination of the three, and finally, performing an intrasubject classification.

The results show that the best model for predicting mental state from EEG data is the LightGBM model without overlapping windows and without applying any feature extraction, with an accuracy of 90.92% and a run time of 14s. Combining EEG, ECG and GSR signals achieves an accuracy of 95.03%, which is a significant improvement over using EEG data alone, but due to the intrusiveness of the Enobio device its use for stress detection cannot be justified. For its use to be feasible, the EEG signal recording device should be wearable.

Contents

List of Figures

List of Tables

CHAPTER 1
Introduction

Stress is a problem that is widespread throughout the world, dubbed by the World Health Organization as the *Health Epidemic of the 21st Century*. Although it is natural to have a certain level of stress due to the challenges we face on a daily basis, stress is not only caused by factors external to the individual, in many cases it is related to the way we interact with the environment and the internal processes involved. In Europe alone, more than 50% of workers and students suffer from stress.

There are positive and negative emotions [24]. Stress is part of the negative ones and if suffered continuously can lead to the appearance of other emotions such as anger or sadness. This set of emotions can be captured through the collection of electroencephalographic or EEG signals from the human brain. In most cases, the procedure of the various studies in this field consists of inducing a calm state in the subject to subsequently make him enter into different emotional states while these contrasting signals are recorded. The major advantage of EEG is that it only records electromagnetic waves from the individual's brain and is therefore non-invasive.

The primary objective of the present Master Thesis is to classify stress and calm brain states from EEG data acquired in an inductive stress experiment and contrast them with other biological signal data such as GSR (EDA) and electrocardiogram (ECG).

Secondary objectives include comparing the performance of different Machine Learning and Deep Learning techniques to corroborate which is the most suitable for studies in this area and a study on the importance of channels to assess the feasibility of using an EEG wearable.

Chapter 2
State of the art

Over the past two decades, multiple studies have been conducted to detect stress in humans using physiological sensors. Some of these earlier studies demonstrate the importance of EDA and ECG signals for stress detection, as in [7, 8, 19]

In 2006, J. Zhai [7] developed a system for assessing stress levels from four physiological signals: electrodermal activity (EDA), skin temperature (ST), blood volume pulse (BVP) and pupil-lar diameter (PD). To classify to resting and stress states the authors used three pattern detection algorithms: SVM, NB and decision trees.

In 2012, Y. Deng [8] demonstrated the importance of feature selection in this type of studies by proposing a method based on Principal Component Analysis (PCA). The authors apply five classification algorithms (SVM, Naïve Bayes, KNN, Linear Discriminant Function and C4.5 induction tree) to evaluate the effectiveness of the feature selection proposed method in accuracy and computational time.

In 2019, R. Martinez et al. [19] designed a system to identify Relaxation Response (RResp), the bodily reaction of an individual when relaxing that positively affects the organism regardless of its emotional state. Analyzing only the GSR [1] signal, the authors identified three levels of RResp (*Low Relaxation Response*, *Medium Relaxation Response* and *High Relaxation Response*) that could be quantified with two calculated features and classified using a decision tree.

Moreover, in recent years other studies have proposed various methods to detect stress from EEG. Kamińska et al. [13] investigated the use of EEG signals to classify the stress level of subjects while using virtual reality (VR). The acquisition protocol consisted of alternating between simulated relaxation and stress scenes while monitoring with an EEG headset. They then compared the performance of a CNN with other Machine Learning methods by stress classification,

[1] Galvanic skin response (GSR) is another term for electrodermal activity (EDA), which measures continuous variations in the electrical characteristics of the skin such as conductance.

obtaining the best result (96.42%) with an SVM considering all brain waves.

In 2022, Phutela et al. [23] proposed a stress classification system using a 4-electrode Muse EEG headband. The experiment consisted of showing four clips, two of stress induction and two of comedy scenes. Once the data were acquired with the headset, they compared two classification models, Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM), achieving an accuracy of 93.17% with the LSTM architecture.

These previous studies on stress detection are the basis on which this work has been inspired. While procedures based on physiological signals such as ECG or GSR have been previously corroborated, the use of EEG signals for emotional detection is a more open and recent field where multiple data acquisition devices and classification models can be applied, since there is no standardized method. While previous studies with EEG focus on stress prediction only with this signal, this project aims to contrast the prediction results with EEG data and other physiological data (ECG and EDA), acquired simultaneously in the same experiment, to assess whether this combination provides a significant benefit that the signals do not provide separately.

CHAPTER 3

Preliminaries

This chapter presents the background necessary to understand the methodology followed in this project.

EEG. The electroencephalogram (EEG) is a neurophysiological examination based on the recording of brain bioelectrical activity by means of special electrodes to study the functioning of the central nervous system from the electrical currents formed in the brain neurons. This procedure allows the diagnosis of alterations in the brain's electrical activity that can lead to diseases such as epilepsy or Alzheimer's disease [12], among others.

Band-pass filter. A band pass filter is a type of filter that allows a specific frequency range of a signal (like the EEG, ECG or GSR) to pass and rejects or attenuates the rest of the frequencies [10].

Frequency bands. Frequency domain analysis has been widely used to interpret the raw recording of EEG signals. Multiple studies have identified five main frequency bands for EEG signals, also called brain rhythms, and have established the correlation between behavior and neural activity in a given part of the brain. Although there is no universal definition of the range of these bands, the following values are generally considered: Delta (0-4 Hz), Theta (4-8 Hz), Alpha (8-14 Hz), Beta (14-30 Hz) and Gamma (30-45 Hz) [21].

Delta waves are related to continuous attention tasks; Theta waves are associated with the state of deep meditation; Alpha waves imply a relaxing state of mind; Beta waves are related to active thinking or anxiety; finally, Gamma waves are used to confirm certain brain diseases.

Machine Learning. Machine learning is a field of artificial intelligence (AI) that comprises the set of techniques that provide systems with the ability to automatically learn and improve from experience [20]. ML allows to perform classification tasks, regression or pattern recognition among other tasks. The fields of application range from medicine, market analysis, robotics or video games.

LightGBM. LightGBM is a fast, distributed, high-performance decision tree-based algorithm supported by the gradient boosting framework, which supports both classification and regression tasks [2]. LightGBM uses a binning algorithm to represent the data in a discrete view (histogram) to find the optimal split point.

The difference with other boosting algorithms is that it splits the tree leaf-wise instead of level-wise. This further reduces the prediction losses and therefore greatly increases the accuracy compared to other similar algorithms.

Another main characteristic of LightGBM is that by using an algorithm based on the histogram, its training speed is higher than the rest of the decision trees, since it groups the values of the continuous characteristics in discrete intervals that accelerate the training procedure.

By replacing continuous values with discrete intervals, memory usage decreases and efficiency increases.

CNN. Convolutional neural networks are a subclass of neural networks that are based on layers of convolutions. Convolution is a mathematical operation that transforms two functions into a third function that represents the magnitude of overlap of both original functions. Like all other networks, they are based on supervised machine learning (i.e. to be used with labeled data). Being a black box algorithm, they are very difficult to interpret [3].

KNN. KNN is a non-parametric lazy learning algorithm that does not build any explicit model. It relies on data similarities and sophisticated distance metrics to generate accurate predictions. In classification tasks, it takes into account the closest observations to a new sample to predict its class [11].

SVM. An SVM is a supervised learning model that represents the sample points in an N-dimensional space, where N is the number of features, to later search for a hyperplane that allows classifying the data points. When it receives new samples, these are classified according to the space to which they belong [9].

CHAPTER 4
Methodological Contribution

To validate the contribution of EEG in stress detection, the following steps are proposed:

- Acquisition of data from volunteers
- Preprocessing of EEG data
- Preprocessing of ECG and GSR data (for later comparison with EEG)
- Data labeling and balancing
- Model building

The following diagram summarizes the last four steps of the methodology followed in this project (see page 29).

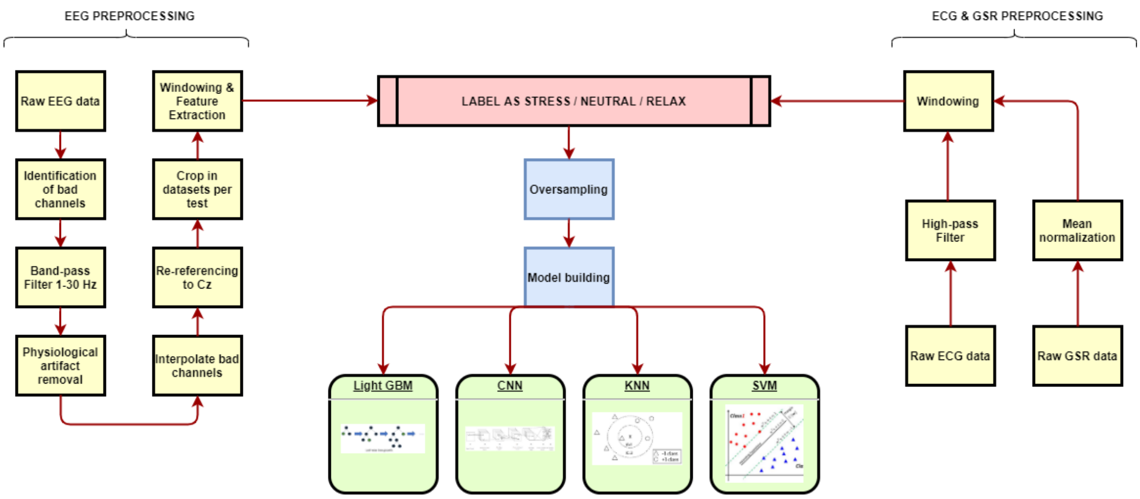


Figure 4.1: Diagram of the methodology

The data collected from all subjects are used to construct the prediction models (multisubject model). This is a decision that favors the availability of a larger amount of data, which are subsequently analyzed in the experimental work performed.

4.1 Data collection

The first part of this project consisted of obtaining data from humans. To that end, an experimental inductive stress procedure is carried out with the aim of eliciting emotional changes in several subjects within a controlled and as realistic as possible laboratory environment. To carry out this experiment, data were collected from six men and four women, all of them students between the ages of 18 and 38.

The acquisition of the participants' data was carried out with the EEG Enobio device from Neuroelectrics and the Biopac MP36 for physiological signals (see Figure 4.2).

(a) EEG Enobio (b) Biopac MP36

Figure 4.2: Signal acquisition devices

Biopac has one set of electrodes for capturing ECG signals and another for GSR/EDA signals. The latter were placed on two fingers of the subjects. For ECG signals, the electrodes were placed following the configuration shown in Figure 4.3.

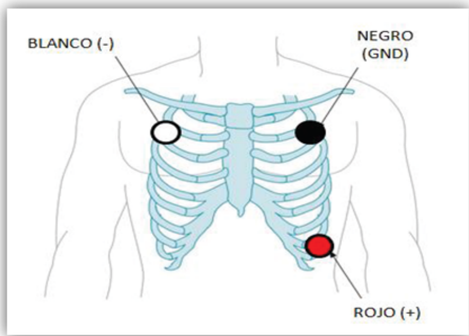


Figure 4.3: Electrode Positioning

The *Biopac Student Lab* was the software used to process these signals, configuring three channels. Channel 1 recorded the ECG signals, Channel 2 the GSR signals and Channel 3 calculated the R-R interval of Channel 1. The sampling frequency was 500 Hz in all three channels.

For EEG signal acquisition with the Enobio device, an EEG25 configuration with 19 standard electrodes was used (see Figure 4.4). Data were recorded with the *NIC2* software with a sampling rate of 500 Hz.

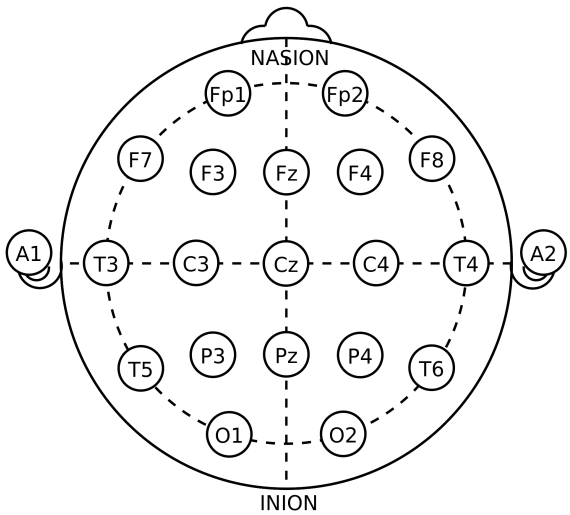


Figure 4.4: EEG standard electrode configuration

The methodology of the experiment was based on the performance of three stress induction tests. Between each test a two-minute relaxation video was shown to the subject. The tests that were performed are as follows.

- **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.

  At the end of the test the subject is shown the solution. Sometimes the solution on paper is more difficult to understand than doing the puzzle itself, which adds 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 an increasing difficulty. Depending on the subject's knowledge of this type of video game, the difficulty was adjusted to challenge the subject.

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. At the end of the experiment, an interview and a questionnaire were conducted to contrast this information with the data gathered by the sensors.

The complete protocol of the experiment is shown in Table 4.1.

Test	Duration (min)	Number of samples (approx.)
Relaxation video 1	2	60.000
3D Puzzle	7	210.000
Relaxation video 2	2	60.000
Mathematical calculations	3	90.000
Relaxation video 3	2	60.000
Video game	4 (approx.)	120.000
Relaxation video 4	2	60.000
Interview & survey	-	-

Table 4.1: Complete experimental procedure

Data acquisition results in two files for each subject:

- An *edf* file which stores EEG data
- A *csv* file which gather ECG and GSR data

Moreover, the questionnaires were recorded on paper and subsequently digitized.

4.2 EEG Data Preprocessing

To perform the EEG data preprocessing, the Python MNE library, a package for visualizing and analyzing human neurophysiological data, was used. There are several methodologies for working with EEG data [15]. The following steps have been chosen for this project, adapted from the manual *Introduction to EEG-preprocessing: Methodological working in imaging neuroscience* [1].

4.2.1 Working with metadata

The first step after importing the raw EEG data is to define the type of acquisition channel. In this project all channels were EEG. Next, the electrode assembly that was used with the Enobio device must be selected, in this case a *standard_1020*, since 19 electrodes were used.

4.2.2 Identification of bad channels

Once the data collection configuration has been defined, a band-pass filter from 1 to 30 Hz is applied. In this way, it is possible to identify which electrodes have noise or no signal and mark them as bad in the initial configuration.

4.2.3 Downsampling

The next step is to apply a 500 Hz downsampling, which reduces the data size. After downsampling, the data has been stored as .csv file.

4.2.4 Additional physiological artifact removal

In this step, an ICA analysis is applied to detect the ocular components and separate them from the brain components. ICA is a common practice that consists of searching for a linear transformation that minimizes the statistical dependence between the components involved in the signal in order to subsequently eliminate artifactual sources from each EEG sample and reconstruct the signals without these components [28].

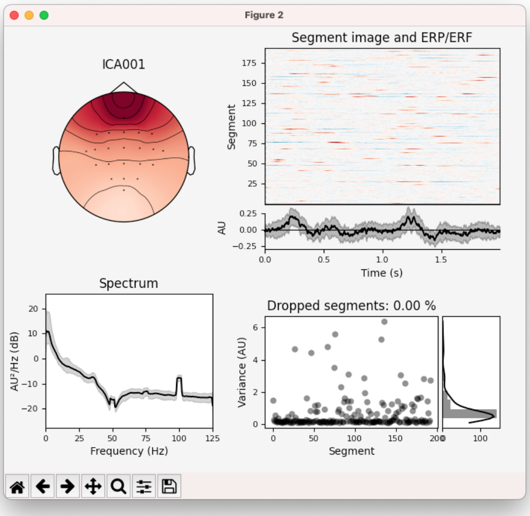


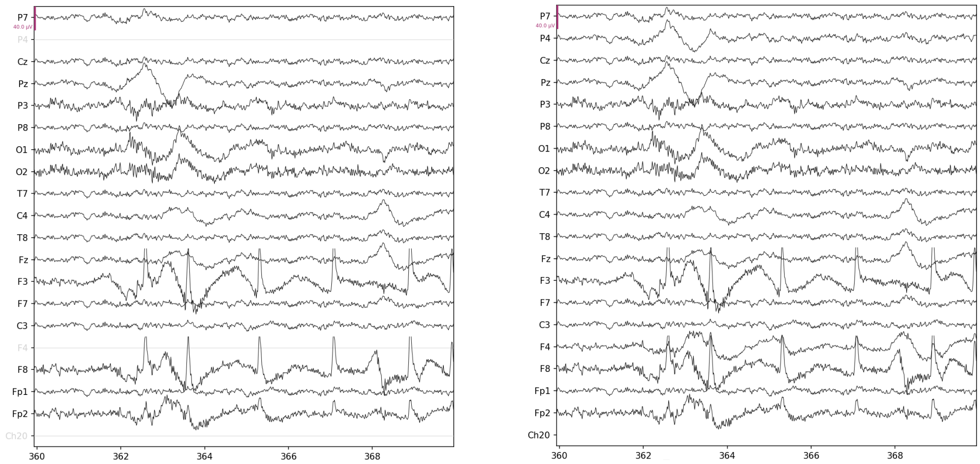
Figure 4.5: Example of ocular artifact in an ICA analysis

Thanks to ICA, eye movements and eye blinks can be removed and the signal corrected. Failure to filter out such artifacts can result in the loss of important information for the detection of the subject's emotional state.

4.2.5 Interpolate bad channels

As this is a multi-subject analysis, it is necessary to maintain the same dimensionality for all subjects (same amount of channels). Therefore, in this case it is preferable to interpolate the bad channels instead of eliminating them, since the bad channels do not have to coincide between subjects.

Channels were interpolated using the spherical spline method [22], which projects the sensor locations onto a unit sphere and interpolates the signal at the bad sensor locations based on the signals from the good locations.



(a) P4 and F4 channels marked as bad (b) P4 and F4 channels interpolated

Figure 4.6: Interpolation of the channels previously marked as bad

In Figure 4.6(a) we observe that channels P4 and F4 have been previously marked as bad, since in this case they did not give any signal. In Figure 4.6(b) these two channels have been reconstructed by interpolating their signal according to the good sensors around them.

4.2.6 Re-reference

In EEG data recording, the most typical references are the vertex electrode (Cz), single or linked earlobes, or the tip of the nose. There is no consensus on which is the most appropriate method, but some studies claim that earlobes introduce more noise than a scalp channel [4].

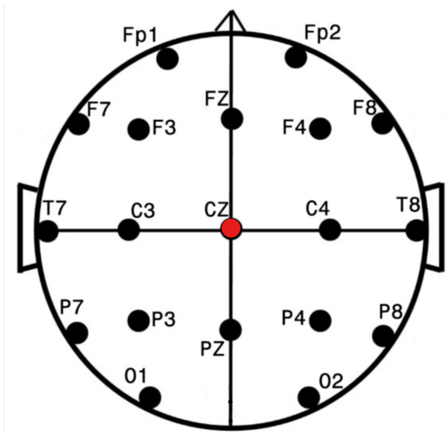


Figure 4.7: Re-reference to Cz

In this project, all channels were re-referenced from the earlobe (GND) to the Cz electrode, since the chosen configuration is symmetrical.

4.2.7 Crop

This step consists in performing a crop of the dataset for each subject test. Thus, seven datasets (three tests and four relaxation videos) are created for each subject, one per step of the protocol (see Table 4.1), which will allow us to easily label the samples later.

4.2.8 Windowing & Feature Extraction

Windowing consists of dividing a signal into different windows with the same length (and, therefore, with the same number of samples) to later apply some transformation and calculate different characteristics that summarize the information of the window. These windows can have overlapping, which means that two consecutive windows share a percentage of their data.

Two approaches have been followed for feature extraction. The first approach followed is similar to the one described in *A Simplified CNN Classification Method for MI-EEG via the Electrode Pairs Signals. Front. Smoke. Neuroscio. 2020* by Xiangmin Lun et al. [16], using only the values of the filtered signals and without performing feature extraction beyond calculating the mean and standard deviations of the chosen windows for each channel.

The second procedure consisted of calculating the Delta, Theta, Alpha, Beta and Gamma bands (also known as brain rhythms) of each window, since they are related to the different mental states, as shown in previous studies such as Somayeh Mohammady's *Wavelet Theory* [21]. As there is no standard definition of the range that defines each of the bands, the ranges defined in the book *Wavelet Theory* were applied:

Frequency band	Range (Hz)
Delta	0 - 4
Theta	4 - 8
Alpha	8 - 14
Beta	14 - 30
Gamma	30 - 45

Table 4.2: Frequency bands

In both cases, windows of 500 samples were selected, so that each window represented 1 second, since, as mentioned above, the sampling frequency is 500 samples per second. Since there is no consensus on whether overlapping is preferable for EEG data, two datasets were created for each approach, one without overlapping and one with an overlapping of 250 samples.

4.3 ECG & GSR Data Preprocessing

For the preprocessing of the ECG signals, a high-pass filter was applied to eliminate noise. In the case of the GSR signals, a mean normalization was applied, since each subject has a different sweating level and the environmental conditions (temperature and humidity) of the laboratory may not coincide between subjects.

Subsequently, as with the EEG signals, two approaches were followed, one without overlapping windows and the other with an overlapping of 250 samples.

These physiological data have been used to make a comparison with the EEG data in the classification results.

4.4 Labeling & Oversampling

Once the data had been preprocessed, labelling was carried out and the classes were checked to see if they were balanced in order to subsequently train the models.

4.4.1 Labeling

In a first stage, labeling aims to consider the problem as a binary one; stress versus relaxation. But as in some samples it was not clear whether the subject was stressed or relaxed, a third class called *neutral* has been added.

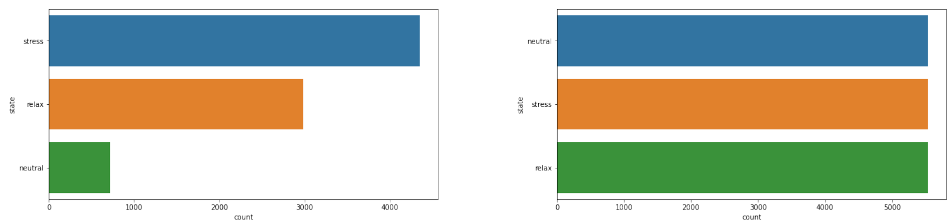
To decide with which state to label each time period, the ECG and GSR data recorded with the Biopac and the interviews conducted with the subjects after the experiment, where each one assessed which tests and which specific moments had caused them more stress, were used as a reference.

Figure 4.8: Labeling with ECG and GSR signals

Figure 4.8 shows a fragment of the captured ECG and GSR signals. As can be seen, the ECG signal stabilizes when the subject enters a relaxed state. At the same time, the GSR signal decreases in that state, indicating a decrease in the subject's sweating, while in the state of stress its tendency is to increase.

4.4.2 Oversampling

Subsequent to labeling, oversampling was applied, since samples in the *stress* class accounted for more than 50% of the samples, as shown in *Figure 4.9(a)*.



(a) Unbalanced samples (b) Balanced samples

Figure 4.9: Oversampling

It is necessary to have the data balanced because at the time of creating the train and test partitions, a class with few samples could have little representation in the train set and, therefore, it would be difficult for the model to classify that class. Oversampling is applied instead of undersampling because the number of samples available is already limited for some models and we do not want to lose information from the majority class.

This balanced the data by replicating the minority class samples and equalized the ratio between the three classes.

4.5 Model building

Four different models were trained on the datasets.

- Light GBM, a gradient boosting framework that uses tree based learning algorithm. This model has the following parameters: learning_rate, n_estimators, criterion, max_depth. Each approach will be initialized with the following values: learning_rate = 0.1, n_estimators = 200, criterion = 'squared_error', max_depth = 15. These parameters were set as an experimental basis and subsequently optimised for each test.
- A 16-layer CNN based on the one proposed by Acharya et al. in the paper *Multi-class Emotion Classification Using EEG Signals* [5].
- A KNN with Grid Search for Hyperparameter Tuning, namely metric, weights and number of neighbors.
- A Radial Basis Function SVM, since RBF kernels are the most widely used form of kernelization and have the best results in most cases [27].

All models have been built with the 80% of the data. The remaining data were used for testing.

CHAPTER 5
Results

To test the hypothesis about the contribution of EEG to stress detection, several experimental scenarios have been carried out:

- EEG data with different feature extraction technique, to select the best set of features.
- EEG data without overlapping and with overlapping of 250 samples, to select the best windowing approach.
- A binary classification of *stress*/*relax* vs the rest, to check if there is a class with dubious labelling.
- A comparison between EEG data, ECG/GSR data and the combination of both, to assess whether EEG data make a significant contribution to prediction.
- An intrasubject classification, to test whether samples from any subject cause conflicts in the classification and whether there is gender bias.

All experiments have been tested with the four differents machine learning techniques (LightGBM, KNN, SVM and CNN). As mentioned in the methodology section, all models have been trained with the same data and the same proportions of train (80%) and test (20%) partitions were used to classify mental state according to stress level. To validate the results, a 5-Fold cross-validation has been applied. Therefore, the data are split into 5 folds and in each iteration one of these folds is used to test the model while the rest are used to train the model.

To evaluate model results the metric used is accuracy in percentages (%):

$$Accuracy = \frac{Number \quad of \quad correct \quad predictions}{Total \quad number \quad of \quad predictions} * 100$$

5.1 Best features

The first experiment was defined to analyze which is the best approach for feature extraction: simple feature extraction (mean and std only), the extraction of the frequency bands and the combination of both methods. In all three cases, the best results for test set prediction were achieved by LightGBM, followed by KNN. With only the mean and std of the samples, LightGBM has an accuracy of 90.92% while KNN decreases to 84%. Training the models with the frequency bands, the results are quite lower, with 76.26% in the best case (LightGBM). Combining the two previous approaches, the accuracy increases again to values close to those of the first method, with a maximum of 89.27% accuracy. This implies that the models give more weight to the mean and std features than to the bands to predict the state.

	LightGBM		CNN		KNN		SVM	
	Train	Test	Train	Test	Train	Test	Train	Test
Mean & std	100.00	90.92	53.42	51.57	100.00	84.00	68.25	63.65
Frequency bands	99.80	76.25	33.66	32.70	100.00	71.00	38.15	36.98
Mean & std + bands	100.00	89.27	53.52	52.02	100.00	80.00	59.92	57.50

Table 5.1: Classification results without overlapping

These results might indicate that frequency bands are not very suitable for predicting stress following the methodology of this experiment.

5.2 Overlapping windows

Since EEG signals are time-continuous data, the effect of applying overlap on the samples had to be tested.

	LightGBM		CNN		KNN		SVM	
	Train	Test	Train	Test	Train	Test	Train	Test
Mean & std	100.00	86.24	49.54	49.10	100.00	78.00	65.23	63.65
Frequency bands	99.23	77.45	34.24	31.73	100.00	72.00	39.86	39.40

Table 5.2: Classification results with overlapping

As the table shows, no improvement is observed in any model and in the case of LightGBM, the accuracy in the test set decreases to 86.24%.

5.3 Binary versus multiclass problem

In the previous predictions, a tendency of the models to misclassify the *neutral* state was observed. For this reason, the three-class classification problem was reduced to a binary stress/relax problem. In the stress vs. relax case, the accuracy of the best model (LightGBM) in the above tests was reduced to 80.65% while the CNN increased to 69.03%. On the other hand, in the relax vs. rest classification, both LightGBM and KNN significantly increased the accuracy.

	LightGBM		CNN		KNN		SVM	
	Train	Test	Train	Test	Train	Test	Train	Test
Stress vs the rest	99.95	80.65	75.24	69.03	100.00	74.00	65.73	63.65
Relax vs the rest	100.00	92.05	49.73	51.08	100.00	85.00	67.89	65.86

Table 5.3: Binary classification results

Analyzing these results it can be seen that samples labeled as *neutral* in case of being misclassified have a tendency to be classified as *stress* while samples labeled as *relax* are easier to classify.

5.4 EEG alone versus EEG with ECG and GSR

The next step was to compare the results of the best model trained only with EEG data, in this case the LightGBM without feature extraction and overlap, with the results obtained from training the same models with ECG and GSR data recorded with Biopac and with the combination of EEG, ECG and GSR data.

	LightGBM		CNN		KNN		SVM	
	Train	Test	Train	Test	Train	Test	Train	Test
EEG	100.00	90.92	53.42	51.57	100.00	84.00	68.25	63.65
ECG+GSR	95.80	87.19	33.80	31.46	100.00	87.00	60.15	58.77
EEG+ECG+GSR	100.00	95.03	54.97	54.34	100.00	87.00	60.15	58.77

Table 5.4: EEG and ECG & GSR classification results

The results show that combining EEG signals with ECG and GSR signals provides a significant improvement in prediction over using EEG signals alone. This increase is more than 4% in the LightGBM model and 3% in the KNN model. On the other hand, combining the signals results in an 8% increase in accuracy compared to using only the physiological signals from the Biopac.

5.5 Personalisation

Personalised or Precision Medicine is a concept that relies on building models based on individual, rather than multi-subject data. Patients are stratified according to their characteristics, risk, prognosis or response to treatment using specialised diagnostic tests. The key idea is to base medical decisions on individual patient characteristics rather than population averages in order to achieve the best possible outcome through personalised treatments [6].

For this reason, an intrasubject classification was performed to analyse whether it is feasible to create personalised models without losing accuracy and to check whether any subject's samples cause conflicts in the classification of the models. For each subject, its sample set was divided into the same training and test ratio as in the previous cases and all four models were trained.

The test results show that the LightGBM model has an average accuracy of 80.53% while the KNN is 75.80%. In general, subjects 1 and 6 have the worst results in all models, which seems to indicate that the samples of these subjects are not as reliable as those of the rest.

The low accuracy of the CNN compared to the other models is due to the fact that being a Deep Learning method it would need a much larger amount of data than those provided in this experiment.

Subject	Gender	LightGBM		CNN		KNN		SVM	
		Train	Test	Train	Test	Train	Test	Train	Test
S0	M	100.00	86.48	33.80	35.61	100.00	76.00	50.24	45.78
S1	M	100.00	66.96	49.14	53.44	100.00	64.00	61.98	59.42
S2	F	100.00	86.01	48.69	55.24	100.00	80.00	72.34	70.28
S3	F	100.00	87.14	50.36	48.56	100.00	85.00	65.15	61.15
S4	M	100.00	74.57	49.53	51.89	100.00	68.00	61.34	57.56
S5	F	100.00	95.69	50.44	48.25	100.00	95.00	84.04	81.67
S6	F	99.80	64.90	49.74	51.02	100.00	64.00	63.12	59.18
S7	M	100.00	88.10	49.66	51.13	100.00	78.00	72.89	69.13
S8	M	100.00	70.02	49.87	50.51	100.00	69.00	57.37	54.62
S9	M	100.00	85.40	49.49	52.02	100.00	79.00	71.05	69.80
Avg.	-	99.98	80.53	48.07	49.77	100.00	75.80	65.95	62.86

Table 5.5: Intrasubject classification results

In addition, we checked for gender bias and as the results show there is no significant difference between males and females.

5.6 Computational time

A comparison between the accuracy with the EEG data of the first approach (mean and std without overlapping) and the execution time for each model shows that LightGBM is undoubtedly the most efficient and fastest model, with an accuracy of 90.92% and an execution time of 14.78s.

	LightGBM	CNN	KNN	SVM
Accuracy	90.92	51.57	84.00	63.65
Time	14.78	204.17	89.84	26.71

Table 5.6: Execution time (s) and accuracy (%) of each model

Even though it is a less complex model than CNN and KNN and with few hyperparameters to tune, it is able to provide the best results among the four models.

5.7 Channels

Finally, it was proposed to carry out an analysis of which channels provide the most information for predicting stress states, quantifying the feature amounts in percentages of all channels (see Table 5.7). The existence of a specific area that is more relevant for prediction would allow the electrode cap to be replaced by a less intrusive portable EEG device.

Importance (%)	Channels
6.0 to 6.8	O1, O2, Pz, T8, F7
5.3 to 5.7	P4, P8, F3, Fz, Fp1, Fp2
4.6 to 5.0	P7, P3, T7, C4, Cz, C4, F4, F8

Table 5.7: Channel analysis

There are two slightly different areas (see Figure 5.1). In magenta are the five most important (values between 6.8 and 6%) and in yellow the six with values between 5.7 and 5.3%. The rest are between 5 and 4.6%.

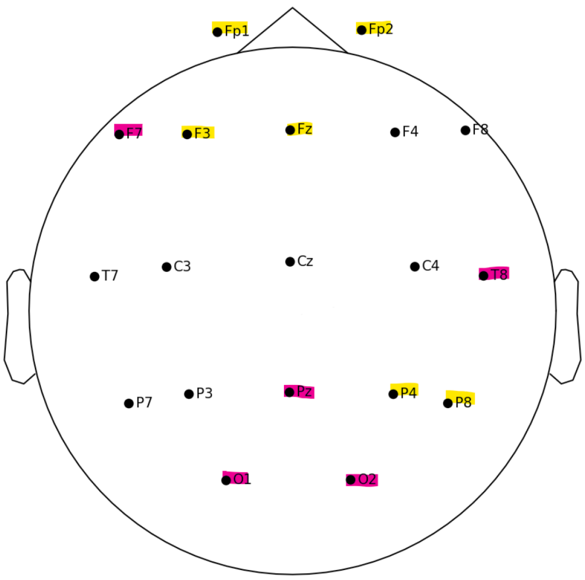


Figure 5.1: Channel importance

Nevertheless, the differences between the percentage values are very small and therefore these results cannot be said to be significant.

5.8 Discussion

The best features are clearly the average and std. Frequency bands reduce the accuracy of all models. Comparing the models, LightGBM and KNN have the best results, with 90.92% and 84% respectively.

Comparing the results with and without overlapping, despite being signals, the models with overlapping provide between 5 and 6% more accuracy.

Analysing the binary classification, the results of relax vs the rest, with an accuracy of 92.05%, compared to stress vs the rest, with an accuracy of 80.75%, seem to indicate that the neutral class tends to be misclassified as stress.

The combination of EEG signals with ECG/GSR, increases the accuracy in the LightGBM model by 5.05% compared to the same model trained with EEG data only and by 8.00% compared to the model trained with ECG/GSR data. The other three models (KNN, CNN and SVM) do not show any improvement and their results are much lower than those of LightGBM.

Regarding the intrasubject classification, the results show that although the average accuracy of the subjects (80.53%) is much lower than the multisubject model (90.92%), in 6 of the 10 subjects the accuracy is close to or even better than the multisubject model.

Comparing execution times, LightGBM is the fastest and highest performing model, with an average of 14s per run.

In summary, the model that according to the results is the most suitable for detecting stress is LightGBM with simple feature extraction (mean and std) and no overlapping.

Analysing the importance of EEG channels, there are no significant differences indicating that some sensors provide more valuable information than others. To corroborate these results, an experiment with a larger number of subjects would have to be conducted.

CHAPTER 6

Conclusions and future work

In this project, various methodologies for stress detection with EEG signals have been contrasted. Using a stress induction experiment, physiological data were acquired from ten subjects. Subsequently, four models have been trained: LightGBM, a 16-layer CNN, a KNN and a SVM.

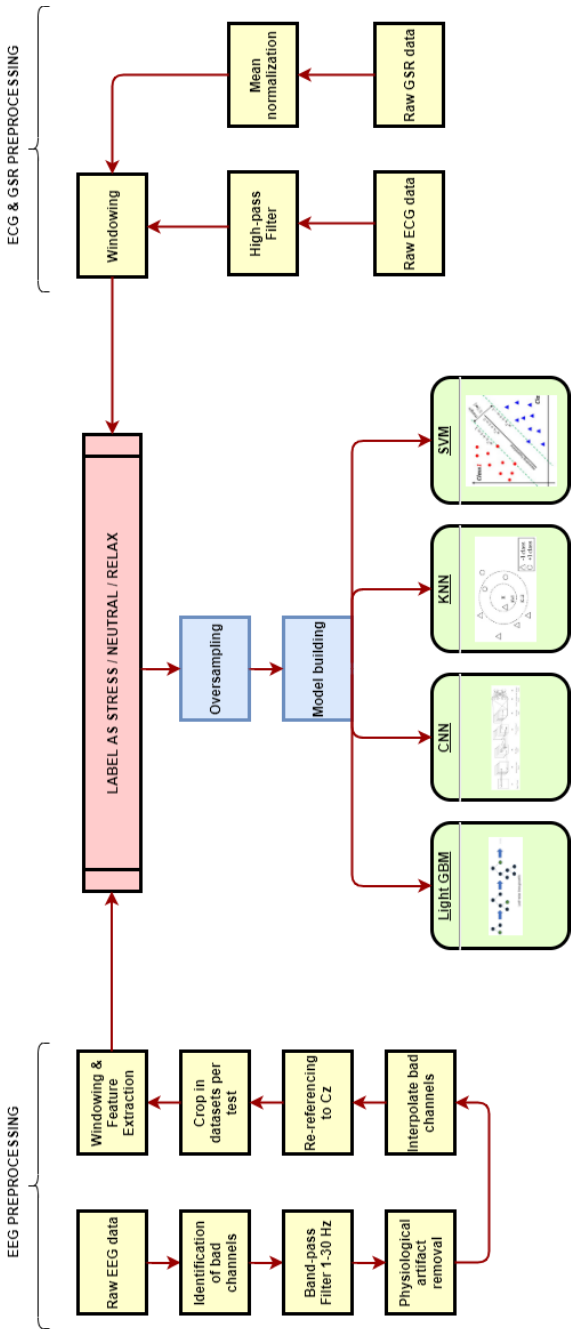
The first hypothesis that had been proposed was whether the inclusion of EEG signals provides any considerable benefit in stress detection versus using only ECG and GSR signals. The results show that EEG signals provide a significant increase in accuracy, between 4-5%, but not enough to justify their use in conjunction with ECG and GSR signals. The major drawback has to do with the issue of intrusiveness. A wearable device, such as a wristband, can be used to record physiological signals, whereas a cap with electrodes is required to capture EEG signals. On the other hand, portable EEG devices are being developed which, unlike the Enobio device, would not be intrusive and could replace the wearable or even be complemented by this type of device.

The second hypothesis that had been raised was which model is the most suitable for the methodology followed by this project. The results clearly show that LightGBM is the most effective, fastest and least complex model. It provides the best results, with an accuracy of 90.92%, only with the means and std of each of the 19 channels, while the second best model, KNN, has an accuracy of 84%. In terms of execution time, LightGBM is much faster than the other models.

In future work, in order to compare the performance of a portable EEG device with a wearable as discussed above, the experiment should be repeated, since in this project the Biopac device was used to record the physiological data and the results could vary drastically with a wearable device.

On the other hand, repeating the experiment with a larger number of subjects would allow the results to be consolidated in order to subsequently write an article.

Diagram of the methodology



Bibliography

[1] José C. García Alanis. Introduction to eeg-preprocessing: Methodological working in imaging neuroscience. (Cited on page 11.)

[2] MJ Bahmani. Understanding lightgbm parameters (and how to tune them). (Cited on page 6.)

[3] IBM Cloud Education. What are convolutional neural networks? (Cited on page 6.)

[4] EEGLAB. Backgound on eeg references. (Cited on page 13.)

[5] D. Acharya et al. Multi-class Emotion Classification Using EEG Signals. *Communications in Computer and Information Science*, vol. 1367. Springer, IACC 2020. (Cited on page 17.)

[6] Holger Fröhlich et al. From hype to reality: data science enabling personalized medicine. *BMC Med*, 2018. (Cited on page 22.)

[7] J. Zhai et al. Stress detection in computer users through non-invasive monitoring of physiological signals. *Biomed. Sci. Instrum.*, 2006. (Cited on page 3.)

[8] Y. Deng et al. Evaluating feature selection for stress identification. *IEEE, International Conference on Information Reuse and Integration*, 2012. (Cited on page 3.)

[9] GeeksforGeeks. Introduction to support vector machines (svm). (Cited on page 6.)

[10] Lawrence P. Huelsman. *Analog Electrical Filters*. Encyclopedia of Physical Science and Technology (Third Edition), 2003. (Cited on page 5.)

[11] IBM. What is the k-nearest neighbors algorithm? (Cited on page 6.)

[12] Jaeseung Jeong. EEG dynamics in patients with Alzheimer's disease. *Clin Neurophysiol*, 2004. (Cited on page 5.)

[13] Dorota Kamińska, Krzysztof Smółka, and Grzegorz Zwoliński. Detection of Mental Stress through EEG Signal in Virtual Reality Environment. *Electronics*, vol. 10(22), 2021. (Cited on page 3.)

[14] Mingu Kang, Siho Shin, Jaehyo Jung, and Youn Tae Kim. Classification of Mental Stress Using CNN-LSTM Algorithms with Electrocardiogram Signals. *Journal of Healthcare Engineering*, vol. 2021, 2021. (Not cited.)

[15] M. Klug and K. Gramann. Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *Eur J Neurosci*, 2021. (Cited on page 11.)

[16] Xiangmin Lun, Zhenglin Yu, Tao Chen, Fang Wang, and Yimin Hou. A Simplified CNN Classification Method for MI-EEG via the Electrode Pairs Signals. *Front. Hum. Neurosci.*, 2020. (Cited on page 14.)

[17] R. Martinez. Diseño de un sistema de detección y clasificación de cambios emocionales basado en el análisis de señales fisiológicas no intrusivas, 2016. (Not cited.)

[18] R. Martinez, 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*, 2017. (Not cited.)

[19] 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. *IEEE Access*, 2019. (Cited on page 3.)

[20] MathWorks. What is machine learning? (Cited on page 5.)

[21] Somayeh Mohammady. *Wavelet Theory*. IntechOpen, 2021. (Cited on pages 5 and 14.)

[22] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, Volume 72, Issue 2, 1989. (Cited on page 12.)

[23] Nishtha Phutela, Devanjali Relan, Goldie Gabrani, Ponnurangam Kumaraguru, and Mesay Samuel. Stress Classification Using Brain Signals Based on LSTM Network. *Computational Intelligence and Neuroscience*, vol. 2022, 2022. (Cited on page 4.)

[24] Elizabeth Scott. How negative emotions affect us. (Cited on page 1.)

[25] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Li Zhenqi, Dan Liao, Xiangmin Xu, and Xinyi Yang. A Review of Emotion Recognition Using Physiological Signals. *Sensors*, 2018. (Not cited.)

[26] Siuly Siuly, Yan Li, and Yanchun Zhang. *EEG Signal Analysis and Classification: Techniques and Applications*. Springer, 2017. (Not cited.)

[27] Sushanth Sreenivasa. Radial basis function (rbf) kernel: The go-to kernel. (Cited on page 17.)

[28] Lisha Sun, Ying 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. (Cited on page 12.)