Universitat de Girona
Escola Politècnica Superior

Treball Final de Màster

Estudi: Màster en Ciència de Dades

Títol: Human stress detection using EEG signals

Document: Resum

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

Contents

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.

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. While previous studies with EEG focus on predicting stress with this signal alone, this project aims to contrast 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.

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.

CHAPTER 2
Methodological Contribution

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 following diagram summarizes the methodology followed in this project.

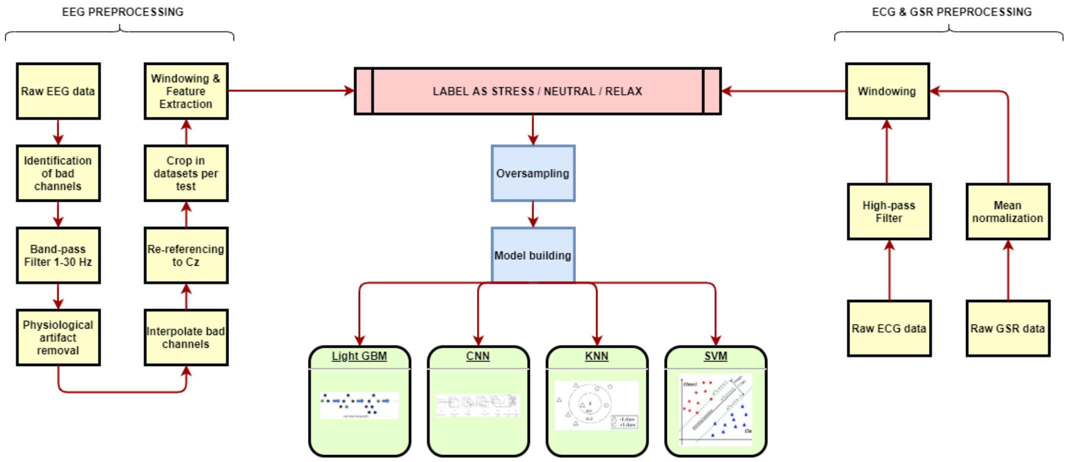

Figure 2.1: Diagram of the methodology

CHAPTER 3

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.

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 (%).

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.

CHAPTER 4

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.