Universidade de Aveiro
2021

Bruno Miguel
Vilela Mendes

Análise do testemunho ocular utilizando sinais de eletroencefalograma

Analysis of eyewitness testimony using electroencephalogram signals

Universidade de Aveiro
2021

Bruno Miguel Vilela Mendes

Análise do testemunho ocular utilizando sinais de eletroencefalograma

Analysis of eyewitness testimony using electroencephalogram signals

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica da Prof. Doutora Ana Maria Perfeito Tomé, Professora associada da Universidade de Aveiro, e do Prof. Doutor Pedro João Bem-Haja Gabriel Ferreira, Técnico Superior da Universidade de Aveiro.

““You catch fear,” Ciri repeated proudly, brushing her ashen fringe from her forehead. “Didn’t you know? Even when something bad happens to you, you have to go straight back to that piece of equipment or you get frightened. And if you’re frightened you’ll be hopeless at the exercise. You mustn’t give up. Geralt said so.””

— Andrzej Sapkowski, in Blood of Elves

o júri / the jury

presidente / president

Professor Doutor Armando José Formoso de Pinho
Professor Catedrático da Universidade de Aveiro

vogais / examiners committee

Doutora Ana Rita Assunção Teixeira
Professora Adjunta da Escola Superior de Educação do Instituto Politécnico de Coimbra (arguente)

Professora Doutora Ana Maria Perfeito Tomé
Professora Associada da Universidade de Aveiro (orientadora)

agradecimentos / acknowledgements

Para começar, gostaria de agradecer à minha orientadora, Professora. Doutora Ana Maria Perfeito Tomé, pelo incansável apoio e acompanhamento prestados durante todo o decorrer deste trabalho. Agradeço também aos Professores Doutores Pedro João Bem-Haja Gabriel Ferreira e Isabel Maria Barbas dos Santos pelo auxílio prestado na vertente de psicologia desta tese.
Quero agradecer ao Instituto de Engenharia Electrónica e Telemática de Aveiro pela oportunidade de realizar o presente trabalho.
Agradeço ainda à minha família pela oportunidade de frequentar a Universidade de Aveiro e pela fé que sempre depositaram em mim.
Por último mas não menos importante, gostaria de agradecer a todos os meus amigos, quer aos de longa data quer aos que vim a conhecer em Aveiro, dando uma palavra especial à minha família de praxe por sempre me terem feito sentir em casa. A todos um forte obrigado pelas vezes que me ajudaram a enfrentar desafios de cariz pessoal ou académico e pelas vezes que levantaram a minha moral.

Palavras Chave

testemunha ocular, reconhecimento facial, interfaces cérebro-computador, eletroencefalograma, extração de características, aprendizagem automática supervisionada, SVM, SVM-RFE, ANOVA.

Resumo

A aplicação de técnicas de Interfaces Cérebro-Computador a testemunhas vitais de um crime pode e provavelmente será uma funcionalidade chave no sistema de justiça.
Características de sinais provenientes de eletroencefalograma foram extraídas com informações sobre o seu domínio (tempo ou frequência), e a sua localização espacial e temporal. Para ambos os domínios, dois modelos de classificação diferentes foram aplicados com vista a selecionar as características mais relevantes: um para classificar, ordenar e selecionar as características mais importantes e outro para eliminar recursivamente a característica menos relevante. O modelo utilizado para classificação foi o Support Vector Machine (linear e não linear).
Outras observações sobre as características selecionadas pelas técnicas aplicadas foram realizadas e discutidas tendo em conta o conhecimento disponível sobre reconhecimento facial.
O presente trabalho fornece um estudo experimental sobre os sinais de eletroencefalograma adquiridos numa experiência na qual foi pedido a um grupo de indivíduos para identificar tanto culpado como distrator, sendo que o culpado estava relacionado a um vídeo de cenário de crime mostrado anteriormente.

Keywords

eyewitness, face recognition, BCI, EEG, feature extraction, supervised machine learning, SVM, SVM-RFE, ANOVA.

Abstract

The application of Brain Computer Interfaces techniques to vital crime witnesses could and probably will be a key feature in the justice system.
Features from the electroencephalogram signals were extracted with information detailing their domain (time or frequency), and their spacial scalp and time placement. For both domains, two different classification pipelines were applied in order to select the most relevant features: one to rank and select the top features and another to recursively eliminate the least relevant feature.
The Support Vector Machine (linear and non-linear) is the classification model included in the pipeline.
Further observations on the selected features by the applied techniques were performed and discussed in relation to the available knowledge about face recognition.
The present work provides an experimental study on the electroencephalogram signals acquired from an experiment in which an array of subjects were asked to identify both culprit and distractor being the culprit related to a previously shown crime scene video.

Contents

List of Figures

List of Tables

Glossary

ANOVA	ANalysis Of VAriance
BCI	Brain Computer Interface
EEG	Electroencephalogram
EOG	Electrooculogram
ERP	Event-Related Potential
FN	False Negatives
FP	False Positives
IIR	Infinite Impulse Response
OOB	Out-Of-Bag
RBF	Radial Basis Function
RAM	Random Access Memory
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine-Recursive Feature Elimination
TN	True Negatives
TP	True Positives

CHAPTER 1

Introduction

1.1 Motivation

Before the DNA analysis entered the court houses, around seventy one percent of the innocent convicted criminals had been condemned by eye witnesses [1]. This is a very troublesome value when we take into account that it is not always possible to obtain DNA based evidences and the testimony from a possible eye witness has a huge importance on the final decision. In some cases, there isn't even a physical evidence at all and an eye witness might be all there is to work with.

However, in most cases, the crime itself is something that takes a few seconds and the witness might not had the time to process what was even happening, or maybe he/she only took a glimpse of the situation and consciously is unable to provide viable information to some investigation that is taking place. This may just be the detail missing in order to get a criminal in jail or stop an innocent from being wrongly accused.

1.2 Context

The human brain is a huge multiprocessing system which receives information from our peripheral system, processes it and controls our actions accordingly. A Brain Computer Interface (BCI) is a system which translates thoughts and provides an interface to communicate with the outside world [2]. The application of such devices to vital witnesses could be a key feature to the justice system because it would, in theory, allow a detective to know if the information he/she is being provided with is actually reliable or if it is just outside noise like fear or anxiety clouding the witness' judgement.

1.3 Objectives

Everyone of us, even if at a subconscious level, is susceptible to influences based on some kind of stereotype that ends up clouding our judgement. But what if we could transcend those stereotypes and know if the witness being questioned actually recognized a criminal face even if he/she is not consciously aware of that?

Through collecting data from Electroencephalogram (EEG) signals, this dissertation aims to provide information about which features are of most importance in the field of face recognition. In order to do that, two different approaches will be used:

- All features will be ranked by a statistical method and the most relevant will be provided to a machine learning model for classification.
- All features will be provided to a machine learning model which will select the most relevant features on its own.

The data collection used for this study was acquired in a previous PhD thesis [3]. Feature extraction was performed on the raw data in order to acquire both temporal and frequency information.

1.4 Structure

This master dissertation is organized in 6 chapters, being one of them the Introduction chapter which was just presented. The remaining 5 chapters are described bellow:

- Chapter 2 presents important concepts related to BCIs. It starts with an overview of the human brain by an anatomical point of view followed by information regarding the analysis of EEG brain signals from a psychologists' point of view and then a detailed description of all the workflow of a BCI. This chapter ends with a brief description of the decision making process regarding the technologies used throughout this present study.
- Chapter 3 starts with the presentation of the dataset that will be used for the study and then details the process of extracting usable information (features) from the EEG signals, both in the time domain and in the frequency domain. It ends with a brief study of the discriminate potential of the features by using an univariate statistical test.
- Chapter 4 starts with some overview of the SVM and SVM-RFE concepts, then details the classification phase from the decision making process to the data analysis and the elaboration of the learning models.
- Chapter 5 presents the results obtained in 4 in a more detailed manner by the uses of tables and graphics.
- Chapter 6 summarizes the solution developed and validates it against other studies in this field and some knowledge provided in 2. It ends with some suggestions for possible developments in future works.

CHAPTER 2

Brain-Computer Interfaces

A BCI is a system that translates the brain activity patterns of a user into messages or commands for a given application [4]. They provide a way to develop interaction between a brain and a computer [2]. BCIs have already contributed to a huge variety of field researches and life applications. In Medicine, brain signals are being used for detection and diagnosis of some disorders or diseases like sleep narcolepsy, brain tumors or dyslexia. Studies are also being made regarding rehabilitation of patients that suffered strokes. With the Internet Of Things being increasingly a now-a-day topic, BCIs also have their place in smart environment with, for example, auto-adjustment control systems. They can also be used to analyze the workload mental fatigue of a person. Neurofeedback is also a promising approach for enhancing brain performance and is currently being explored in the ambit of the educational field of research. Last but not least, entertainment and gaming applications have a lot of market offer for BCI from simulators to Virtual Reality (VR) games [5].

2.1 Human Brain

The basic units of our nervous system are the neurons which have a body cell, dendrites with receiving function and an axon with conductive function. Neurons are seen as cells specialized in communication either with other neurons or with other kinds of cells [6]. The communications between them are performed through synapses, which are electrical or electrochemical signal junctions. A given nervous impulse is transmitted to the bottom terminals of the axon who later releases a neurotransmitter (excitatory or inhibitory) in the synaptic gap - the space between two linked neurons - which will be received by the next neuron's dendrites.

The nervous system can be divided in two main parts [7] [8]:

- Central Nervous System (CNS) consisting of the brain and spinal cord.
- Peripheral Nervous System (PNS) consisting of the spinal and the cranial nerves.

The brain is the nervous' system main organ. It can be divided in three parts [7] [8]:

- Cerebrum, largest part in the brain, divided in left and right hemispheres.

- Cerebellum, responsible for coordination, posture and balance.
- Brainsteam, responsible for basic functions such as breathing and heart rate and connects the cerebrum and cerebellum to the spinal cord.

As for the cerebrum, it can be divided in two hemispheres, left and right, in which one is usually responsible for the opposite side of the body. The cerebrum can also be divided in four lobes as represented in figure 2.1, each one with specific tasks but none of them can work on their own. The four lobes are [7] [8]:

- Frontal lobe (personality, emotional behavior, intelligence, concentration, judgement, self awareness, speech and body movement).
- Parietal lobe (interprets hearing and language, spatial and visual perception).
- Occipital lobe (interprets vision).
- Temporal lobe (understanding language, memory and organization).

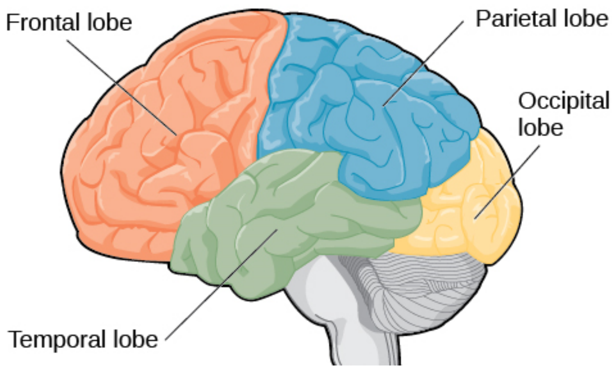

Figure 2.1: The lobes of the brain from a subject looking left. Adapted from [9].

Memory itself is a complex process that can be decomposed into three phases: encoding, storing and recalling. Different brain areas deal with different types of memories (sensations, visuals, hearing). A new face, for instance, when observed is interpreted in the occipital lobe and temporarily stored on the prefrontal cortex where it will remain for some minutes. Doing an analogy with computer systems, a new data is stored in the Random Access Memory (RAM) where it can be accessed in an efficient way. After some time, the memory of this new face may be copied in the form of neuronal activity patterns and stored in the hippocampus of the temporal lobe. Memories associated to strong emotions will form stronger patterns with stronger synapses and can form indelible memories. Taking back the previous analogy, some sets of data will be saved in the hard drive where they can be stored indefinitely. There is also another kind of memory called skill memory, processed by the cerebellum which stores learned procedures like playing an instrument or riding a bike [10].

2.2 Brain Signal Acquisition

The acquisition of brain signals can be performed in three different ways [2]:

- Invasive: penetrating micro electrodes into the brain's dura matter.
- Semi-invasive: electrodes are placed on the scalp but not in the gray matter.

- Non-invasive: electrodes are placed on the scalp, without surgery.

The non-invasive techniques are the most used in research environments due to not provoking any kind of damage to the brain tissues. Among these techniques, the following three stand out [2]:

- Electroencephalogram: EEG signals correspond to the electric potential difference or voltage generated by the neuron activity inside the brain. This potential difference occur mostly due to the sodium and potassium ions going in and out of the cells [11].
- Magnetoencephalogram: records magnetic fields produced by the neural activity. Because the cranium does not cause any attenuation of the magnetic signals, this technique produces very accurate results nevertheless the equipment require a lot of maintenance costs and cares.
- Functional Magnetic Resonance imaging: records changes in the blood flow through the magnetic properties of oxygenated blood. This is the technique with the most spacial resolution but lacks in time resolution.

The EEG is the most used brain signal acquisition technique in BCI environments. It may not be the most accurate technique since the cranium blocks a lot of electric energy but it is the most cheap and practical one since it's non-invasive, lightweight and portable. Forgetting the usage of brain signal acquisition for health reasons, it makes little sense to submit a scientific study's participant to a surgery or spending a high budget in magnetoencephalogram equipment and maintenance.

2.2.1 Electroencephalogram

An EEG is basically a way to write down the brain's electric activity, in other words, the electrical impulses of the neurons while communicating with each other. It's also a technique with a very high time resolution which proves useful when there is a need to study short time windows like, for example, in this study where the main information is within the first second post stimulus [11].

The brain signals are recorded through metal electrodes placed on the scalp. Those electrodes usually positioned according to the standard 10-20 electrode placement system described in [12]. Figure 2.2 shows the disposition of the electrodes in this type of system. From this model, others appeared in which the number of electrodes increased as the distance between them decreased or vice-versa. Figure 2.3 shows the headset that was used to record the data for the present experience. This one has a total of 34 electrodes.

When analyzing EEG signals, one usually tries to study the signals' frequency bands or the behavior of the signal around the occurrence of a given stimulus. According to [14], a stimulus is the prime independent variable of a psychological experiment. It can be considered a change in the present environment. When performing a psychological experiment, there is usually one stimulus per each trial. Figure 2.4 shows a diagram of what a trial with duration t_{trial} can look like. In the present study, the stimulus will be the appearance of a face in a screen in front of a participant.

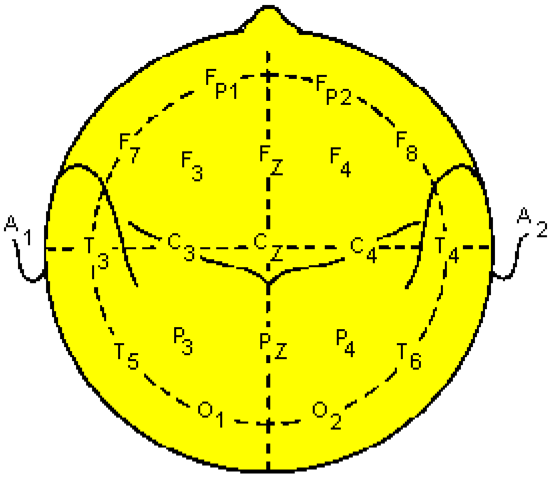

Figure 2.2: Scheme of the electrode disposition in a 10-20 system. The face is turned upwards. Adapted from [13].

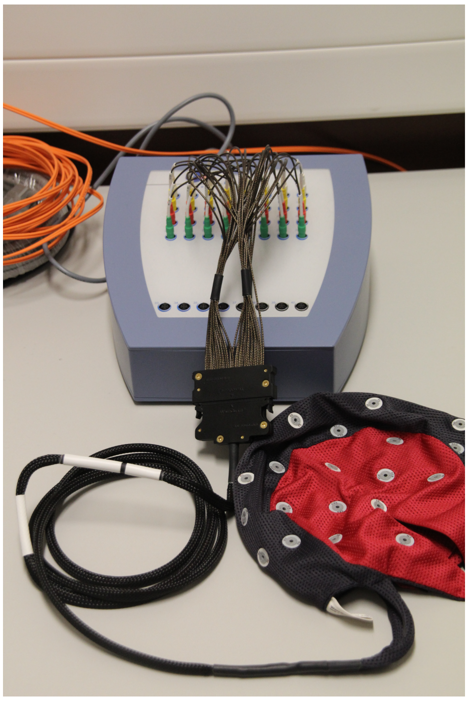

Figure 2.3: The headset (red and black) used to record the data for the present study. Photo taken in the Department of Education and Psychology at University of Aveiro.

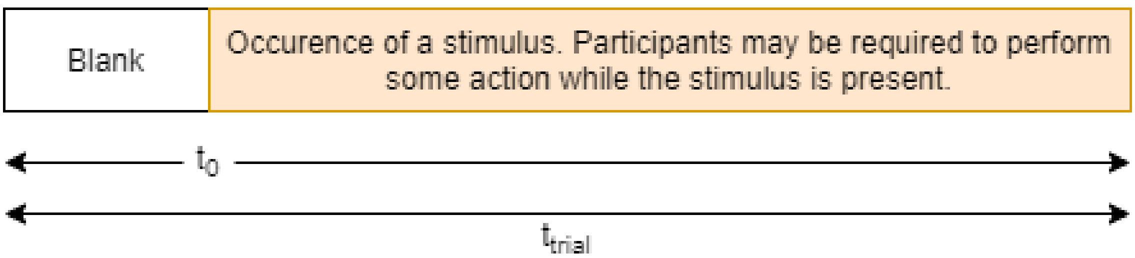

Figure 2.4: Diagram of a psychological experiment's trial. The time instant of the occurrence of the stimulus is marked as t_0.

2.2.1.1 *Frequency-domain characteristics*

One way to look at EEG data is to study its rhythmic activities. These activities can then be divided into frequency bands and studied separately. The brain frequency data can be divided in the following frequencies [11]:

- Delta band [1, 4) Hz: with slow rhythm and high amplitude, delta waves are usually present during sleep. Stronger in the right hemisphere, their source is located in the thalamus.
- Theta band [4, 8) Hz: can be recorded all over the cortex and correlates with mental operations such as focused attention, learning or memory recalls.
- Alpha band [8, 13) Hz: usually associated with relaxation and closed eyes, their band power is almost suppressed during activity with eyes opened. Alpha suppression can be a signature of states of mental activity of engagement.
- Beta band [13, 30) Hz: generated in posterior and frontal regions. An higher beta power can correlate to active, busy or anxious thinking.
- Gamma band [30, 50) Hz: these waves have only recently been discovered and studies about them are still very fresh.

Comparing with the time domain, components in frequency domain are easier to identify but require more computational resources [15]. Figure 2.5 shows the average shape of these brain waves in a normal adult.

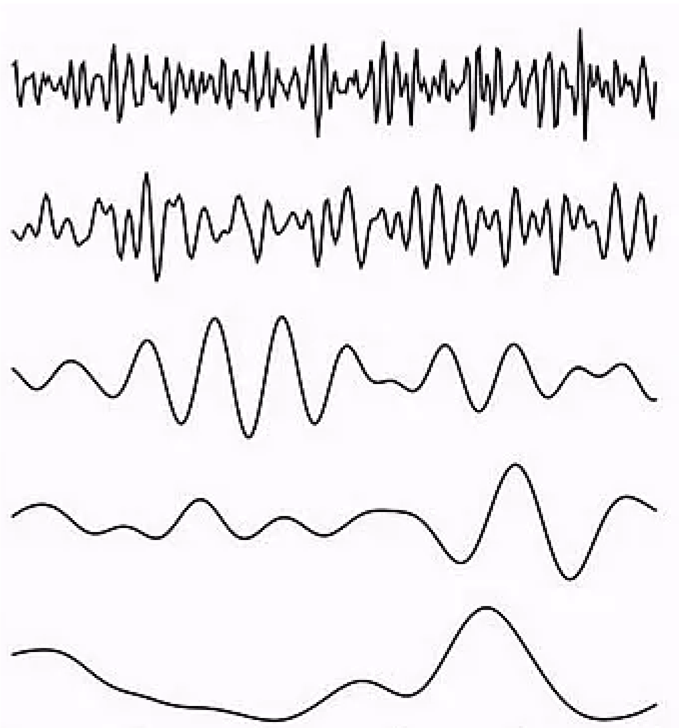

Figure 2.5: Brain waves in a normal adult with a time interval of one second. From top to bottom: gamma, beta, alpha, theta and delta. Adapted from [16].

2.2.1.2 Time-domain characteristics

Another way to look at EEG data is through a time perspective point of view. This analysis is computationally simple but the conclusions may be limited. It requires the data being recorded in an epoch-based paradigm which involves the occurrence of stimulus all separated by an equal period of time [17]. A signal from an epoch can be divided into three different components: a short time window before the stimulus, a time point identifying the occurrence of the stimulus and a long time window after the occurrence of the stimulus. Each epoch can be studied as a single occurrence, making a single-trial study, or averaging similar epochs related to a particular stimulus to provide a signal wave which is less affected by noise.

The data is read in sequence of the cognitive operations that are happening and mapped on peaks. These peaks are the dominant way of data gathering and when they are extracted from an epoch-averaged signal they are traditionally called Event-Related Potential (ERP) [18]. They are basically brain neuron activities stimulated by a given response. An ERP waveform is a series of positive and negative voltage deflections related to a set of underlying components [19]. In an ERP waveform, the positive and negative components are identified by their temporal occurrence and are represented, respectively, with the initial P or N followed by either their latency in milliseconds or their count number [2] [15]. For example, a P200 is the positive-going peak that occurs around 200 milliseconds after a given stimulus and a P2 is the second positive-going peak in a wave after a given stimulus. They may or be not refer to the same ERP component. Note that a P-ERP component doesn't actually need to have a positive value of amplitude as long as it is a local maximum. The same logic goes for N-ERP components. Figure 2.6 shows a N170 that happens to have a positive value but it is still a negative-going peak.

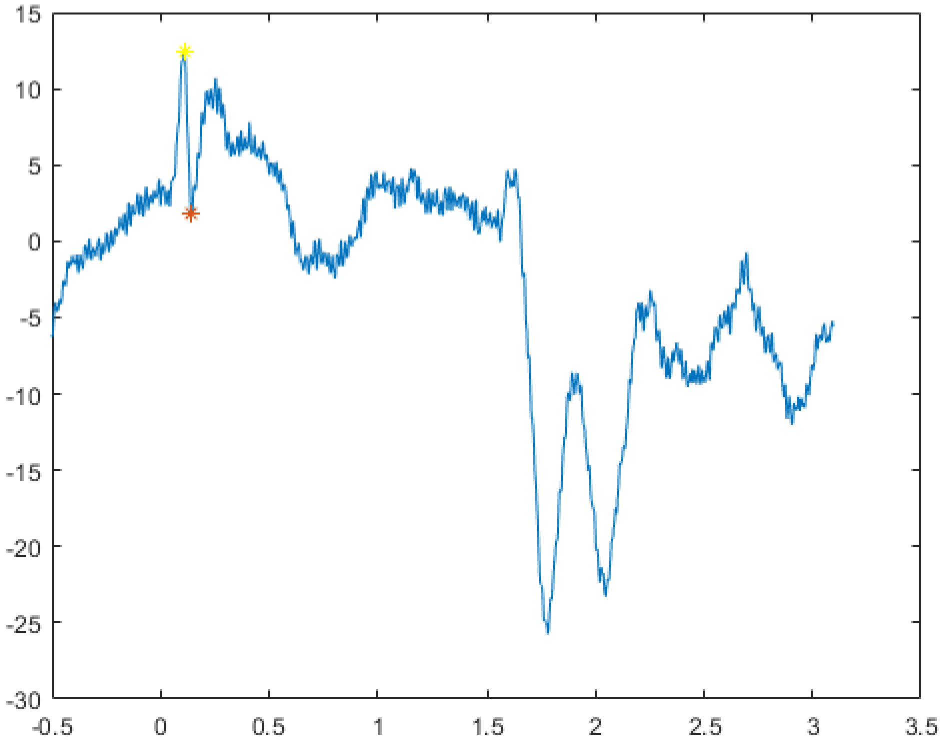

Figure 2.6: An EEG signal from an occipital channel with a P100 marked in yellow and a N170 marked in red. The x-axis represents the time in seconds and the instant of time zero marks the occurrence of a stimulus.

2.3 BCI workflow

A BCI can be summarized in two essential work phases: signal processing with feature extraction and machine learning. Either implicit or explicit, a preprocessing phase also occurs right after the EEG data extraction. Figure 2.7 represents the basic workflow of a BCI and the next subsections detail some concepts about each phase. Note that the trigger is what was called before as the stimulus.

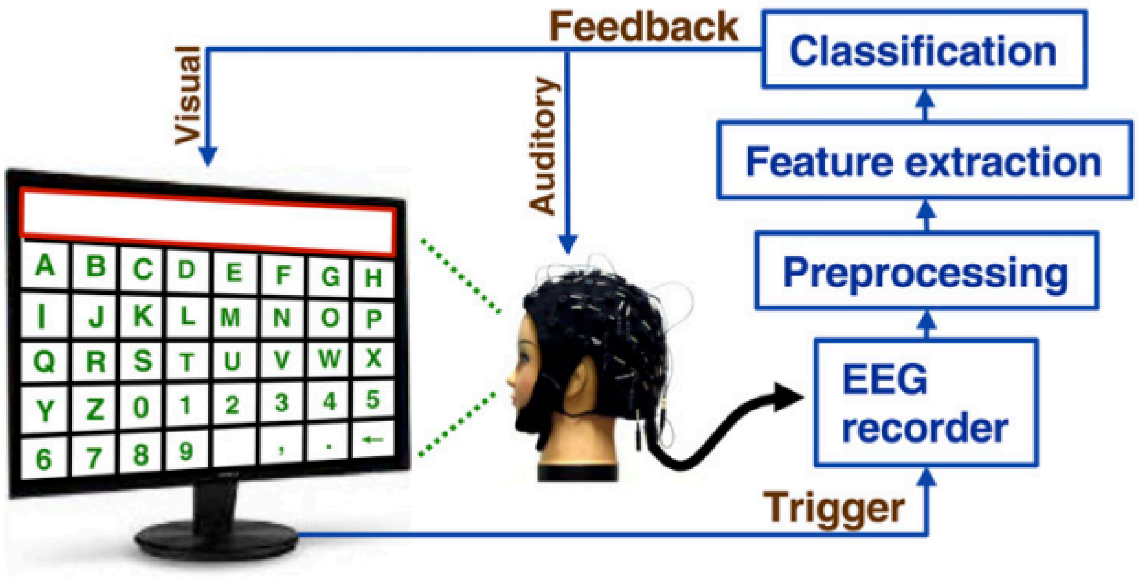

Figure 2.7: Typical workflow of a BCI. Adapted from [20].

2.3.1 Preprocessing

In most cases, this phase is transparent to the person dealing with EEG data extraction since most systems already perform some of the tasks related to it by default.

Preprocessing an EEG signal means eliminating noise data or artifacts, data that has no usefulness for the problem at hands, and in most cases that data is known to be: eye-blinking, electrocardiograms or other internal or external disturbing effects [18]. That is the reason why during an EEG record session, the Electrooculogram (EOG) is also recorded so later it can be used to remove ocular artifacts from the signal. The mastoid channels also recorded are usually used as reference channels.

2.3.2 Feature extraction

Once the EEG signals are acquired, they need to be prepared so that later features can be extracted. The two most common types of features in EEG are time point and frequency band power features related respectively to the time domain and the frequency domain mentioned above.

Time point features are usually extracted after some signal processing, notably a band-pass filter and down sampling and the result are the ERP waveforms mentioned before [4]. Although the concept of an ERP usually refers to a data point extracted from the mean EEG signal, a BCI tries to extract it from a single trial so that it can try to quickly adapt itself to the user. For this reason and throughout this document, the concept of ERP component will be used for either a peak in a single trial or a peak in a mean EEG signal acquired from all the related trials.

The band power features however can be computed in lots of different ways [21] and, being more complex than simple time points, they may also provide an amount of information greater that simple time points [4].

2.3.3 Univariate analysis

This is a topic that some may include at the feature extraction step or in the beginning of the classification step. Before heading to the classification step, a study of the acquired features may come in handy, specially when we are talking about a very large number of features. It is not viable to feed a machine learning algorithm with an extremely large number of features not only because it will get very slow but also because many of these features may have irrelevant degrees of importance.

There are a lot of different techniques to understand each feature independently. Most general cases include histograms, box plots, density plots, correlation matrices and scatterplot matrices [22]. In the specific subject of brain waves, concepts like ANOVA and F-test arise [23] [24]. According to [25], ANalysis Of VAriance (ANOVA) can determine whether the means of three or more groups of features are different and uses F-tests to statistically test that equality. F-tests are named after its test statistic, F-statistic, which is simply a ratio of two variances. Variance represents how far the data are scattered from the mean. This process is quite helpful in determining which are the most relevant features.

2.3.4 Classification

After gathering the data and selecting the features, the final phase begins. Although not always the case, most BCI work with supervised machine learning for classification problems. Supervised machine learning is where you have input, or independent, variables and an output, or dependent, variable and you use an algorithm to learn the mapping function from the input to the output. It is called supervised because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process [26]. Classification is the process of predicting the class of given data points [27]. Basically, an user's action or decision is recorded and the system tries to label or cluster it. In classification problems, this label is usually represented as a class, a non continuous variable. For example, the face presented is either a culprit or a distractor.

The number of classification algorithms is huge. The Support Vector Machine (SVM), while not recent, is still very reliable. It projects the input space to high-dimensional space such that data can be separable. The main goal is to choose the optimal separating hyperplane so that the distance between two data points from different classes can be maximized. An algorithm that has become quite popular in the last few years is the Convolution Neural Network, associated to Deep Learning. It is a type of Neural Networks (inspired from biological neural systems, the inputs are called neurons which are connected with weights) that arranges neurons in three dimensions: width, height and depth. [2] A Convolution Neural Network may also do a processing of the data on its own before performing classification and some may say that it joins both the feature extraction and the classification phases together. Having proven to be quite useful in a lot of subjects, deep learning techniques have not yet convinced in the

BCI field [4]. Another option that is becoming popular is the Random Forest. The Random Forest algorithm is composed of different decision trees, each with the same nodes, but using different data that leads to different leaves. It merges the decisions of multiple decision trees in order to find an answer, which represents the average of all these decision trees [28].

2.3.4.1 Model training and testing

After choosing and setting up a machine learning model, it needs to go through its training phase. As it was referred before, a training phase for a supervised machine learning model requires a dataset with both independent and dependent variables which are provided to the model so it can learn. It is similar to a teacher providing examples to his/her students so they can can train themselves and learn how to solve similar problems.

When the training phase is complete, there is need to evaluate the classification model generated. It is obviously not recommended to use the data we used to build the model to evaluate it. Grabbing the previous analogy, a teacher cannot really evaluate if the students understand the subject taught if he/she evaluates them with the exact examples that they trained in class or in homework. Instead, some similar exercises are provided with small changes but the core subject remains the same.

The most common way to test a model is the holdout method. The purpose of holdout evaluation is to test a model on different data than it was trained on. The dataset is randomly divided into at least two subsets:

- Training set, used to build the model.
- Test set, or unseen data, used to evaluate the performance of a model.

The holdout approach is useful for its speed and simplicity but the results may not always be accurate, specially when the amount of data is limited [29].

Another option is the cross-validation which usually provides more trustworthy results than the holdout model, specially for few data. This procedure has a single parameter called k that refers to the number of groups that a given data sample has to be split into. The general steps for this method are listed below [30]:

1. Shuffle the dataset.
2. Split the dataset into k groups.
3. For each group:
    a) Take the group as a test data set.
    b) Take the remaining as training data set.
    c) Fit a model on the training set and evaluate it on the test set.
    d) Retain the evaluation score.
4. Summarize the skill of the model using all the scores.

The Random Forest method also provides a useful score for validating the model which is the Out-Of-Bag (OOB) score. Usually, each tree in the forest makes use of around two-thirds of the provided training dataset and the rest are called OOB samples which are used for validation. Each tree has its own OOB score and the final result is an average from all the trees' outcome [31].

2.4 Choosing the technology

Before proceeding to the practical part of this dissertation, there was a need to decide what would be the best environment to work. Three programming languages were analyzed as potential candidates: MATLAB, Python and R, for two main phases of work: Signal processing and machine learning. R was excluded firstly due to the learning curve it would need and, despite being somewhat better than MATLAB in the analytical field, it lacks the years of experience of the Mathworks product.

2.4.1 Signal Processing

Python has very diversified applications and has lots of advantages: user friendly, open source and a vast community support (even though it is a relatively recent language). Mainly because of its open source nature, Python is becoming more popular each day in the professional world [32]. It would be a good choice was it not for the years of experience MATLAB has with signal processing. Not only that but MATLAB is still the best option to deal with anything that can be represented as a numeric feature matrix [32]. And if all this wasn't enough, there is the EEGLAB. EEGLAB is an interactive MATLAB toolbox for processing continuous and event-related EEG and other electrophysiological data. EEGLAB also provides an interactive graphic user interface allowing users to flexibly and interactively process their high-density EEG and other dynamic brain data [33]. The only downside of MATLAB is the cost of license but with the University of Aveiro providing a license to all its students, this downside just got deleted. So, for the feature extraction process that deals directly with signal processing, the choice did fall on MATLAB.

2.4.2 Machine Learning

Despite MATLAB being chosen for the first phase, when the problem comes to machine learning, Python is still the most popular language and the reasons are countless. Python comes with a huge amount of useful libraries such as scikit-learn, which is also great for data mining and analysis. As it was said before, Python is a very friendly user language and being so every programmer can easily get start handling Python. When compared to MATLAB, Python code is often more compact and easier to read, all of it is open source and it offers a wider set of choices in graphics packages and toolsets. It may loose a little when it comes to documentation since the available at Mathworks is really good but because Python is open source, a huge amount of people work on it and the information available online makes up for it [32]. So the choice for this last phase fell on Python.

CHAPTER 3

Feature Extraction

The first step in dealing with BCI is the feature extraction. According to [34], feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing.

This chapter starts with a general overview on the dataset to be used further on. Before proceeding with the extraction step, a downsampling will take place since the original sampling rate was very high and frequencies above the beta band are of no interest to this work. Finally, time domain and frequency domain features are extracted from every recorded EEG signals.

3.1 The Dataset

The existing database was the result of an experiment performed in [3] at University of Aveiro. A total of eight theft videos of twenty seconds were displayed to the participants in which the culprit was presented in frontal view during four seconds and not in frontal view the rest of the time. After each video, the EEG recording starts and a lineup with ten cycles of six faces including five distractors and a culprit in random order was presented to the participant. All images were in gray-scale, emotionally neutral and the distractors' face were rated as similar to the culprit's face. Each face would be displayed for one and a half seconds and its appearance is the stimulus considered for the experiment.

The EEG data was recorded from 34 electrodes mounted on a waveguard cap according to the 10-20 system at a sampling rate of 2048 Hz. The channel locations are displayed in figure 3.1. The EOG was also recorded to serve as reference for ocular correction algorithms. The ocular artifacts in the EEG signals were then eliminated using an eye-movement correction algorithm and the resulting data were re-referenced to the average of the left and right mastoid electrodes [3].

There are a total of 29 participants in the present dataset but data from participant number 16 was not considered for the present study.

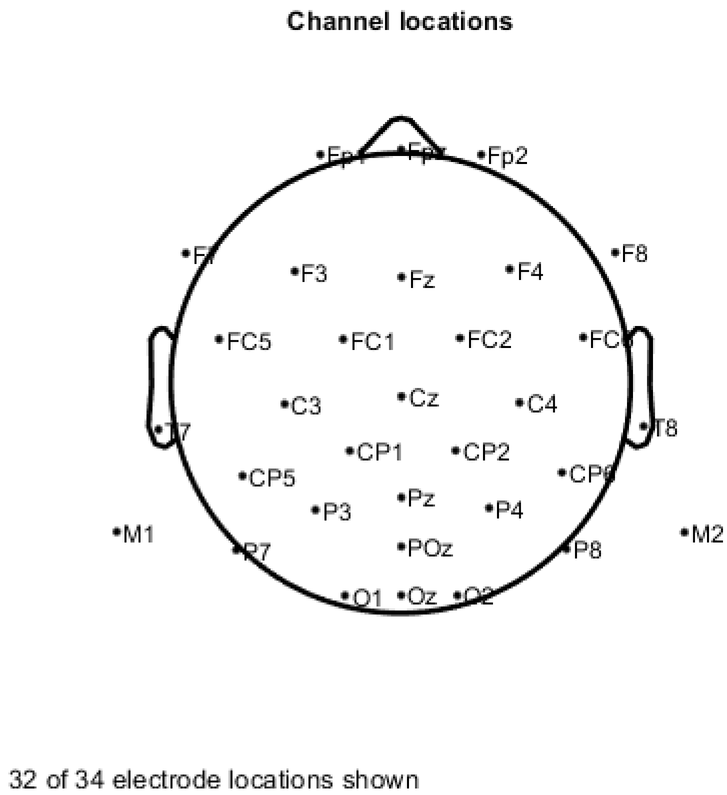

Figure 3.1: Channel locations in the used system. The face is upwards. All but the eye channels are represented.

3.2 Downsampling signals

First thing to do was to downsample the data from a 2048 sampling rate to 512 sampling rate first because the original sampling rate was very excessive and took a lot of computer resources and second, frequencies way higher than the beta band are considered noise in the present case of study. It is very important not to loose the exact time when the face appeared for the subject so a lowpass filter with a zero-phase strategy was applied. Therefore, the data is filtered twice, one in each direction in the time axis, and the time delay that results from the filtering process is canceled out so no time information is lost in the process [35]. Figure 3.2 shows details a strategy used by this type of filter.

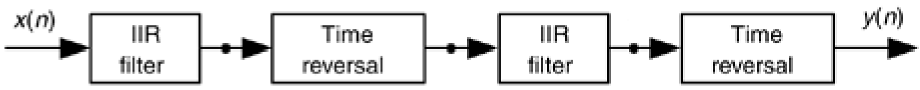

Figure 3.2: A strategy used by a zero-phase filter. Adapted from [35].

Next, each participant's file was processed as follows: the epochs when the participant did not answer or when he/she gave two answers were discarded and the signals were splitted according to four different classes: target and right answer, target and wrong answer, distractor and right answer, and distractor and wrong answer.

As explained in section 2, the delta waves usually presented in during sleep time and therefore are not relevant to the present goal. Theta and beta waves being related to focused attention, memory recall and anxious thinking are objects of interest [11]. As for the alpha waves, studies have been made concerning the asymmetry of this particular wave in questions of emotion and motivation [36].

Frequency analysis of the signal was performed using MATLAB tools. The spectrogram was applied for the signal divided into segments, multiplied by an Hamming window of 0.25s seconds with an overlap 0.125s. The analysis is calculated for the range [4, 30]Hz with a resolution of 0.1Hz. Such Short Time Fourier Transform was achieved and the power values for the signal are available in a 2D array, the number of columns being related with the length of the input signal and the number of rows with the length of the vector of frequencies. The figure 3.3 illustrates the outcome of the analysis of one segment starting one second before the stimulus's presentation.

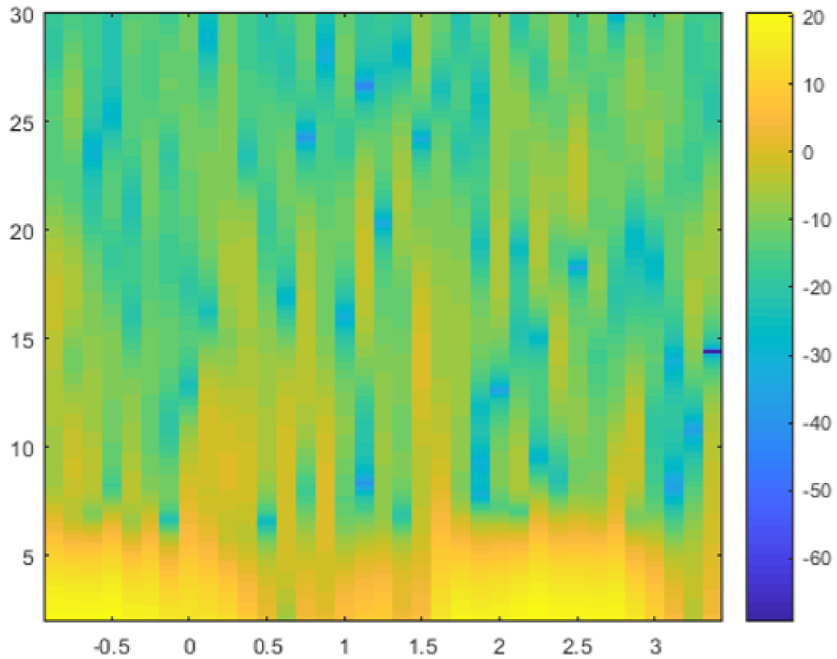

Figure 3.3: Power spectrum in the range [4 30]Hz of a segment of signal (channel FC1) starting 1s before the stimulus. The x-axis represents the time in seconds.

Further processing of this data comprises calculation of the frequency contents in the three bands of the EEG signal: theta, alpha and beta. The energy in theta, alpha and beta band were calculated adding the energy values belonging to the [4, 8)Hz, [8, 13)Hz and [13, 30)Hz respectively. The figure 3.4 represents the information of the analysis of the signals of all channels for a single participant.

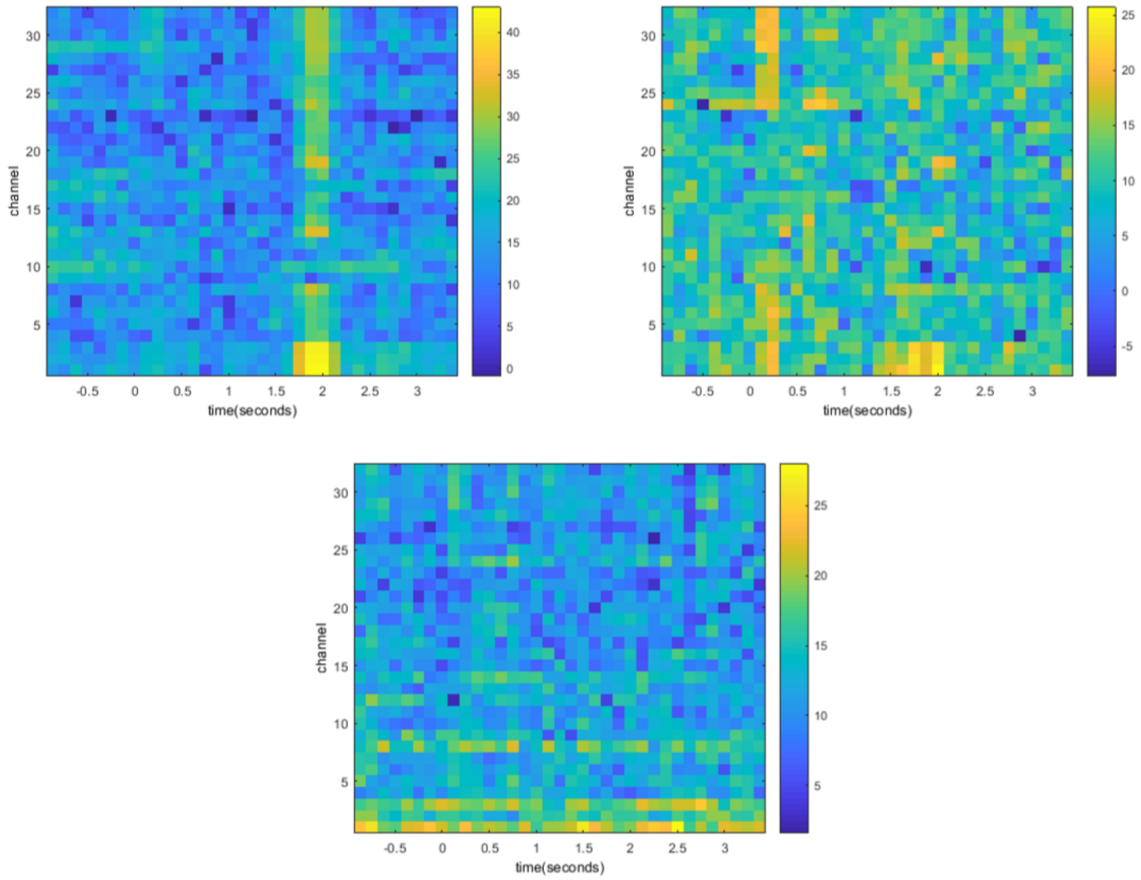

Figure 3.4: Energy in the characteristic bands theta,alpha and beta (left to right). The vertical coordinates of the image represent the index of the 32 EEG signals (all except for the eye channels.

3.4 Time domain analysis

Time domain analysis comprises the detection of peaks (positive or negative) in pre-defined time intervals after the presentation of the stimulus. The peaks are the ERPs components or events. According to the literature [37] [38] [3], the most relevant ERPs components in subconscious face recognition are the P100 (80ms to 120ms) which is mostly related to any visual evoked potential, the N170 (150ms to 190ms) associated with face processing and the P300 (300ms to 600ms) which relates to the occurrence of rare events. An example used frequently when speaking about P300 is the oddball paradigm and it is detailed in [39]. Study [38] also takes into account the P200 (180ms to 220ms).

Two different approaches were taken:

- the ERPs were extracted from the mean signals estimated for the four different conditions;
- the ERPs were extracted on each single trial.

It's usual to use the ERPs extracted from the mean signals to perform population studies over some subject [40]. Figure 3.5 shows the P100 and N170 extracted from the mean signal of one subject in a specific condition. The plots are colour coded accordingly to the colorbar, where red represents the largest positive value and blue represents the largest negative value.

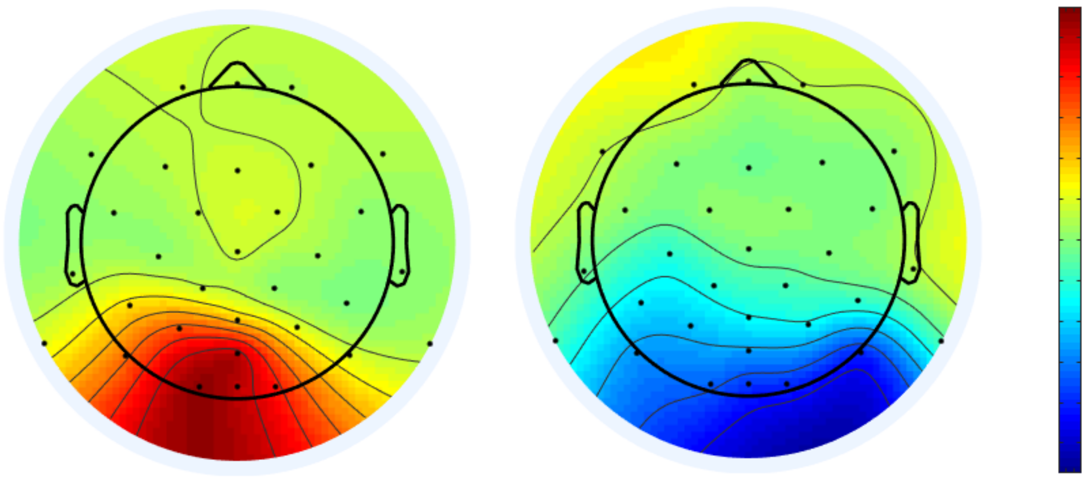

Figure 3.5: Example of a P100 of condition distractor+right (left) and N170 of condition target+right (right) of a single subject acquired from mean signals on the respective ERP interval.

However, in order to match the frequency data for future fusion of both time and frequency and in order to increase the number of samples for the machine learning models, the present project will work with the ERPs extracted from the single trial signals.

For single trial analysis, the signal should be smoothed, therefore the signal is filtered before peak detection. It is important to keep following the zero-phase strategy to preserve the time reference of the stimulus in the output of the filter. In the end, no phase-shift was applied and the filter did not impose any temporal offsets or distortion in the signal. The Infinite Impulse Response (IIR) filter of the block diagram is a 4th passband filter between 2 and 12 Hz designed using the Butterworth approach. The magnitude frequency response of the filter is shown by the black line in the figure 3.6. The global attenuation (in dB) of the zero-phase filtering strategy is the double of the attenuation of the filter as shown by the red dashed line in the figure. Figure 3.7 illustrates the input and the output of the filtering system.

Logically, a visual ERP like a P100 is most likely to occur in the occipital region, or near it (see 2.1). However, in this present work, all the channels were taken into consideration and information related to all four ERP components referred before was extracted. For a single channel and a single ERP, the process to extract information is as follow:

1. Identify if it is a positive or negative ERP component (e.g. positive for P100).
2. If positive/negative, search for the highest/lowest peak in the respective time interval (e.g. for P100, search for the highest peak from 80ms to 120ms).
3. Save both amplitude and latency values.

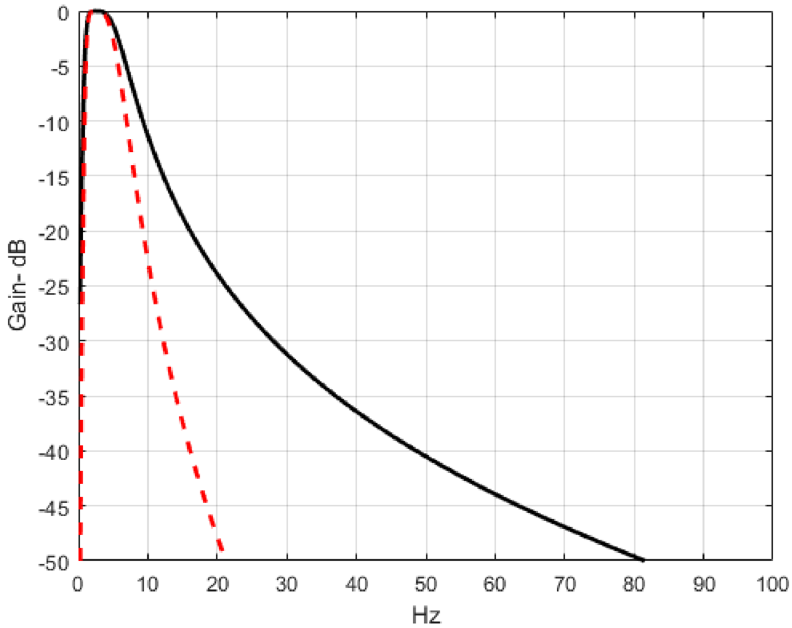

Figure 3.6: IIR filter using the Butterworth approach. The black line represents the magnitude frequency response of the filter and the red dashed line is the magnitude frequency response of the zero phase filtering scheme.

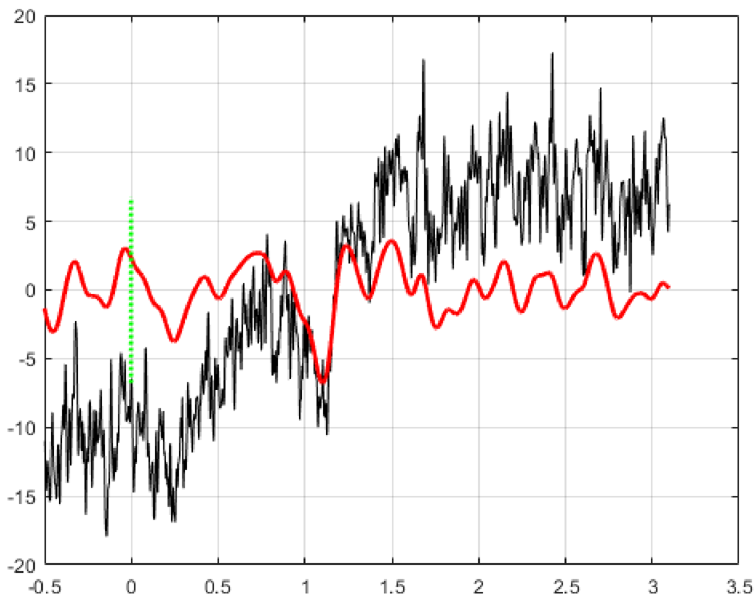

Figure 3.7: Input (black) and output (red) of the zero-phase filtering system when applied to a signal. This signal is from channel FC1. The time is represented on the x-axis in seconds.

3.5 Feature Vectors

From the classifiers point of view, their entries will be a feature vector. According to [41], a feature vector is used to represent numeric or symbolic characteristics, called features, of an object in a mathematical, easily analyzable way. The present work will deal with two kinds of feature vectors: the ones related to the frequency features and the ones related to the time features.

3.5.1 Frequency domain features

The features extracted in 3.3 were summarized in one table per participant with each row representing a feature vector, a single trial with all features represented in the columns. For each trial, the information available was as follows:

- 3 frequency bands of interest: theta, alpha and beta
- for each frequency band, 27 instants of time being the first 3 pre-stimulus and the last 23 post-stimulus.
- for each instant of time, 30 channels were taken into account (all but the eye and the mastoid).

This makes a total of 2430 features per trial which can be concatenated into a vector $\mathbf{x}$ with dimension $D = 3 \times 27 \times 30$.

3.5.2 Time domain features

The features extracted in 3.4 were summarized in the same way as the frequency data except that now there are just a total of 240 time features related with the four single trial ERP components also mentioned in 3.4: P100, N170, P200 and P300. This number represents the multiplication of the following values:

- 4 amplitude values related to the four accounted single trial ERPs.
- 4 latency values related to the four accounted single trial ERPs.
- 30 EEG channels (all but the eye and the mastoid channels).

Like the frequency ones, these features can also be concatenated into a vector $\mathbf{x}$ with dimension $D = 4 \times 4 \times 30$.

3.6 Conclusion

During this chapter, the dataset on which the upcoming classification will take place was presented and the process of signal acquisition was briefly described. The work started with a downsampling of the EEG signals due to their high value of sampling rate. Soon after, the epochs with useless information were discarded and the remaining epochs were splitted into four conditions. The signals were then analysed and features were extracted in two different perspectives: time domain and frequency domain. Both feature vectors are detailed at the end of the chapter.

CHAPTER 4

Classification

In the previous chapter, several data files were generated from the EEG signals with either time and frequency domain characteristics. As mentioned in 3.2, the single trial signals were splitted according to four conditions. These conditions will be analysed in pairs so the problem at hands will always fall to a binary classification problem.

This chapter starts with some overview of the SVM concept, the machine learning model chosen for this work as well as the SVM-RFE concept, its feature selection algorithm. Then the classification phase from the decision making process to the data analysis and the elaboration of the learning models are detailed. The results will be summarizes in the next chapter.

4.1 SVM and Feature Selection

4.1.1 SVM

Support Vector Machines (or just SVMs) were developed in the 90s and are still very popular nowadays in solving supervised learning problems. They are motivated through Statistical Learning Theory (SLT). Solving a supervised learning problem in SLT can be briefly described as follows: given a set of training labeled data and a loss function that measures the error when comparing with the real value of the label, the goal is to find the model parameters that minimize the expectation of the error on new data [42].

SVMs tackle this problem by trying to find a hyperplane that can distinctly classifies each sample. In an N-dimensional space, hyperplane is a flat subspace of dimension N-1 and the N depends on the number of features. In other others, it is a decision boundary. Figure 4.1 shows two different classes of data points (yellow and red) being separated by a hyperplane (blue line). Each data point is treated as a vector and its components are the the features or independent variables. Obviously, a lot of hyperplanes may provide a valid solution to the training dataset. The used criteria is to choose the one with the highest margin. The margin is measured by the distance from the hyperplane to the closest data point in each class. These data points are called support vectors and they influence both the position and orientation of the hyperplane. These concepts are illustrated in figure 4.2 [43]. The present examples show only two independent variable in the x-axis.

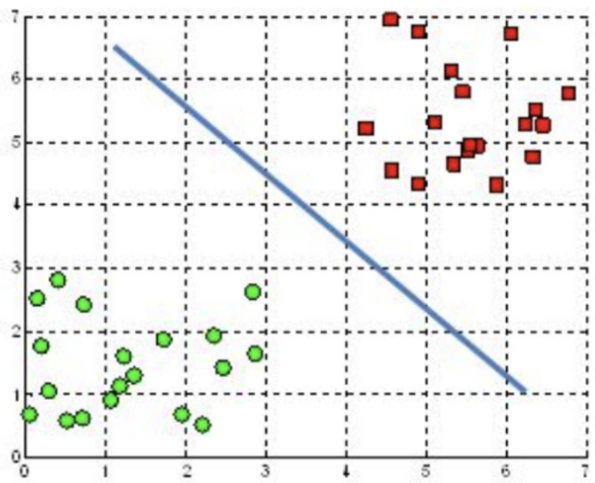

Figure 4.1: a hyperplane (blue line) separates two classes (red squares and green circles). Adapted from [43]

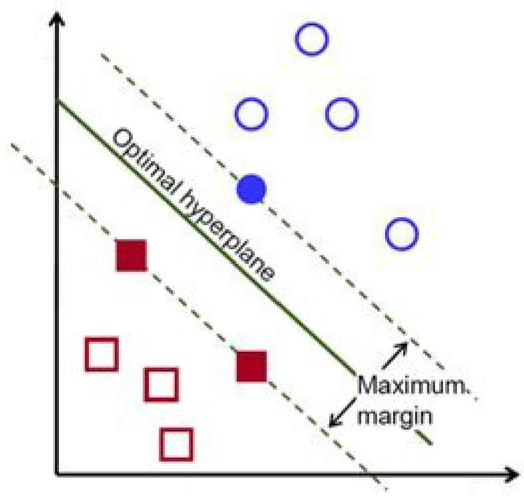

Figure 4.2: The choice of the optimal hyperplane. The colorful points are the support vectors. Adapted from [43]

The problem becomes more complex when it is not possible to draw a plane that is able to separate all classes in the dataset or, in other words, when the data is non-linearly separable. There are two different ways to deal with this situation [43]: the first is to allow the SVM to tolerate a few miss classifications and try to maximize the global accuracy. For that, a degree of tolerance can be defined, usually represented by C. The bigger the value of C, the lesser the tolerance for mistakes; the second is to use a kernel trick by mapping the data into a higher dimensional space and try to find a hyperplane in that dimension. Different kernels map data in different ways. Figure 4.3 represents a dataset in a 2-dimensional space being mapped into a 3-dimensional space in order to find a valid hyperplane.

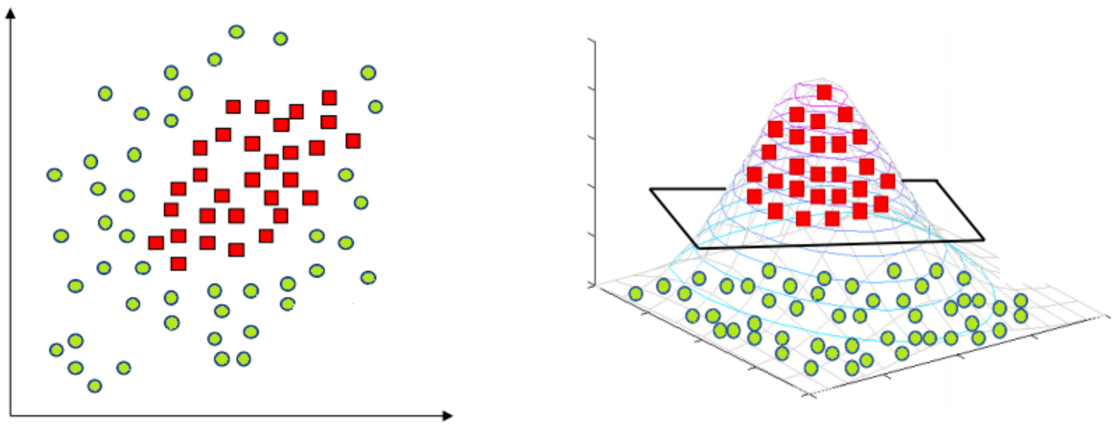

Figure 4.3: The original dataset (left) is mapped into a higher dimension space (right) in order find an optimal hyperplane (black square). Adapted from [43]

Looking the the SVM from a more mathematical perspective, when defined one needs to specify the kernel it will use.

Being $\mathbf{x}$ and $\mathbf{z}$ two feature vectors, the linear kernel is basically a simple dot product [44]:

$$K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T\mathbf{z} \tag{4.1}$$

The Radial Basis Function (RBF) kernel determines a non linear decision surface defined by the following equation where γ is the parameter gamma which must be greater than zero [44]. In this case, a mapping operation is performed simultaneously with the dot product of the mapped data.

$$K(\mathbf{x}, \mathbf{z}) = \phi^T(\mathbf{x})\phi(\mathbf{z}) = \exp(-\gamma \left\|\mathbf{x} - \mathbf{z}\right\|^2) \tag{4.2}$$

To adapt the SVM model, the following dual problem should be solved.

$$\min_{\alpha_i} \sum_{i=1}^{1} \alpha_i - \frac{1}{2}\sum_{i=1}^{1}\sum_{j=1}^{1} \alpha_i\alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}), \tag{4.3}$$

$$\text{subject to: } 0 \leq \alpha_i \leq C, \text{ for all } i$$

$$\sum_{i=1}^{1} \alpha_i y_i = 0$$

The outcome of the optimization are the Lagrangian values α.

The parameter C is defined by the user. When working with normalized data, C is usually set as equal to 1. The result of the learning process are the α values and there is one α for each pair $(\mathbf{x}_i, y_i)$ of the training set. The $\mathbf{x}_i$ with $\alpha_i \neq 0$ are the so called support vectors. Note that when the problem is linearly separable and the kernel is linear, those values lie on two hyperplanes that define the margin of the classifier (see figure 4.2).

In order to classify a new vector $\mathbf{z}$, the decision rule would be:

$$g(\mathbf{z}) = \sum_i \alpha_i y_i K(\mathbf{x_i}, \mathbf{z}) + b \Rightarrow \begin{array}{ll} g(\mathbf{z}) > 0 & y = 1 \\ g(\mathbf{z}) < 0 & y = -1 \end{array} \tag{4.4}$$

Note that when the problem is linearly separable and the kernel is linear, the substitution of $K()$ as defined by eq. 4.1 leads to the calculation of the weight values. They are calculated with the Lagrangians α in the following way:

$$\mathbf{w} = \sum_{i=1}^{K} \alpha_i y_i \mathbf{x_i} \tag{4.5}$$

In that case, the decision rule can be:

$$g(\mathbf{z}) = \mathbf{w}^T \mathbf{z} + b \tag{4.6}$$

Therefor, the entries of the vector $\mathbf{w}$ can be used to eliminate features in vector $\mathbf{z}$. Note that if an entry in $\mathbf{w}$ is zero, the corresponding entry of $\mathbf{z}$ does not contribute to $g(\mathbf{z})$. With all this, linear SVM has its own way to eliminate the less important features. However, as its needs the $\mathbf{w}$ values in order to do it [45], the ranking cannot be done when using a radial basis kernel function or any other kernel function that does not allow to calculate the $\mathbf{w}$ like the linear kernel does [46].

4.1.2 SVM-RFE

Support Vector Machine-Recursive Feature Elimination (SVM-RFE) consists of a wrapped method of feature selection based on the SVM classification algorithm. As explained in **svm**, the SVM classifies the data by dividing it using a hyperplane. The models performance is evaluated by cross-validation and the chosen measure for its evaluation is then calculated. With the SVM-RFE in place, the next step will be to eliminate recursively the less important features. The number of features to be eliminated is passed as an entry value. The process then repeats itself.

The recursive feature elimination method implemented [47] stops when it reaches the number defined as the minimum features to select. The model's performance is measured at each step by cross-validation and after it's finished, it returns information on how many and which features were used when the highest measure value was achieved. Despite being a robust method, the SVM-RFE doesn't take into account the correlation between the features as it is not a statistical method and its worst drawback is the huge amount of time that this process consumes [48]. SVM-RFE has been successfully used in, for example, a brain cognitive study to identify relevant scalp areas [49].

4.2 Model evaluation

As referred before in 2.3.4.1, the classification model must be evaluated with labelled data not included in the training phase. With the outcomes of the model it is possible to create the so called confusion matrix. A confusion matrix is a table constituted by four basics counts that will later be used to compute the measures by which the model can be evaluated. These four counters are [50]:

- True Positives (TP), the number of correctly recognized target class examples
- True Negatives (TN), the number of correctly recognized examples that do not belong to the target class.
- False Positives (FP), the number of incorrectly examples assigned to the target class.
- False Negatives (FN), the number of examples that were not recognizes as the target class.

The table 4.1 is represented the confusion matrix of this work's present situation.

Class \Predicted as	Target	Distractor
Target	TP	FN
Distractor	FP	TN

Table 4.1: Scheme of a confusion matrix.

With the entries of this matrix is possible to calculate all the measures needed to evaluate a machine learning model. The most common when dealing with a balanced dataset is the accuracy which measures the overall effectiveness of a classifier. The formula to calculate the accuracy is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{4.7}$$

However, when dealing with unbalanced dataset, this measure won't provide a viable result because it measures both classes with the same weight. Imagine the case where one given dataset has 10% of the target class and the rest are distractors. If a model classifies every class as distractor, it will result in an accuracy value of 0.9 which may seem really good but it is misleading because in reality the model fails to identify every single target class. So, in this case, other types of measures are required. The one used in this work is the balanced accuracy. The balanced accuracy [50] [51] avoids inflated performance estimates on unbalanced datasets and can be calculated by averaging both the sensitivity and the specificity values. Sensitivity, or recall, measures the effectiveness of the classifier in identifying target labels and specificity does the same but for distractor labels. The formula for these measures are shown bellow:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4.8}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{4.9}$$

$$\text{Balanced Accuracy} = \frac{Sensitivity + Specificity}{2} \tag{4.10}$$

Looking again to the case example mentioned before, now the model would be evaluated with a sensitivity of 0 because it did not label correctly any target and a specificity of 1 because it labeled correctly every distractor. These values will lead to a balanced accuracy of 0.5. More details about these or other types of measures not used in this work can be found in [50].

4.3 Classification Pipelines

The study followed a single subject based analysis as most BCI usually have a training phase for each individual participant [4]. This analysis increased the number of universes but left a reduced the number of samples for each individual classifier to work with.

Having few trials, specially if the dataset had been balanced before, it was decided to use the SVM for the classification since it presents good performance in smaller datasets [52]. Due to the high number of features, it became necessary to select the most important ones. Two approaches were performed in order to achieve this goal:

- The features were ranked using the ANOVA technique.
- The SVM itself, through recursive feature elimination, keeps removing the less important feature(s) and evaluation its performance with the remaining ones.

Both models were evaluated by a cross-validation with $k = 5$ [53]. The preferential measures chosen in order to evaluate the models' performance were the accuracy score [54] for the balanced data and the balanced accuracy score [51] for the unbalanced data.

4.3.1 Feature Ranking

In order to obtain a rank of the total number of features, an univariate analysis was performed using the ANOVA method. The ANOVA is a statistical test that can be used to analyze differences between two or more groups of data. It usually uses the f-test score and makes a single, overall decision as to whether a significant difference is present among sample means. On the particular case of having only two different groups, like the present case which has target and distractor, it instead uses what is called the t-test [55].

To implement the feature ranking, the class SelectKBest [56] was used. The discriminating power of each feature is evaluated individually and assigned a value which will be used for ranking purposes. The features are assumed to be approximately normally distributed [57] so this method is likely to present poor results when dealing with unbalanced datasets. After each feature being tested, they are assigned a rank value and it is that value that will be used to rank the features by their importance. The more they differ between the two classes, the more important they are considered.

After being ranked, the top features were selected and normalized before being provided to an SVM model. Figure 4.4 shows the largest f-values of the ANOVA test calculated for frequency domain features of the signals of one participant.

Data normalization was performed by a Standard Scaler [58] which standardizes the features by removing the mean and scaling it to the unit variance. Being x_i the feature ith,

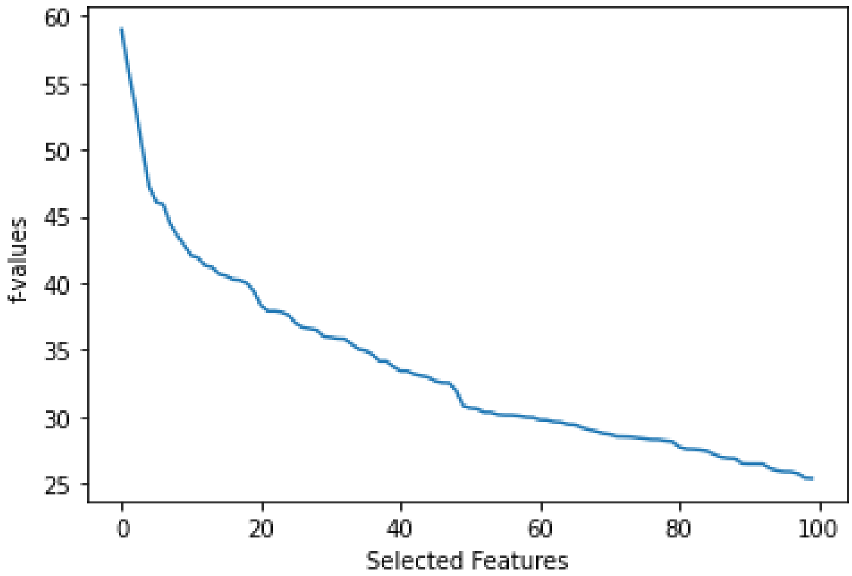

Figure 4.4: F-values for the top 100 features from the frequency domain in one participant.

its normalization is calculated as follows:

$$z_i = \frac{x_i - u}{s} \tag{4.11}$$

where u and s are respectively the mean and the standard deviation of x_i. Both linear and RBF kernels were used.

A complete diagram describing all the implementation steps for the classification with feature ranking is displayed in figure 4.5.

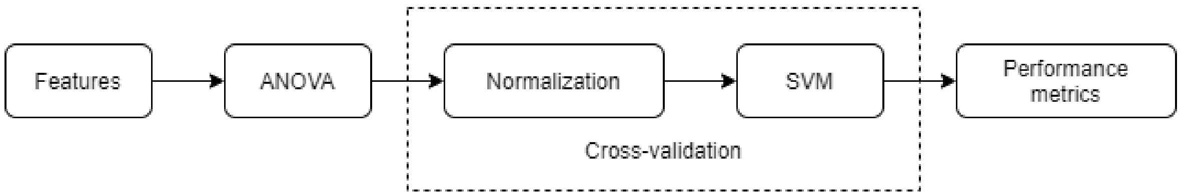

Figure 4.5: Classification using feature ranking.

4.3.2 Feature Elimination

On a second approach to the feature selection step, the SVM was free to choose which feature it wanted to use with the SVM-RFE and then it would evaluate its choice using a linear kernel since it does not support a RBF kernel (see the end of 4.1.1). Figure 4.6 shows a diagram describing the implementation steps for the classification based on feature elimination.

The usage of the SVM-RFE can return a graphic like the one presented on figure 4.7 which shows the result of the measure the user chose for evaluating the model on the yy axis and the number of features selected on the xx axis. It also returns that measure's best score and the number of features required for that score. It is also possible to know which features were used for the classification.

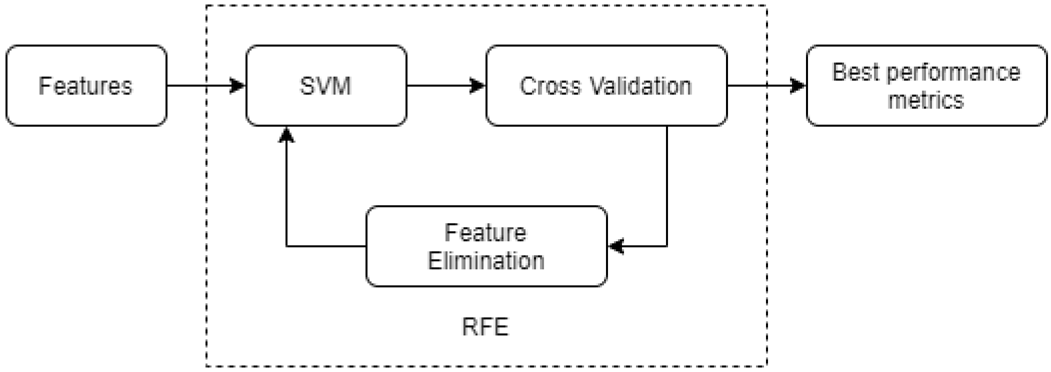

Figure 4.6: Classification using feature elimination.

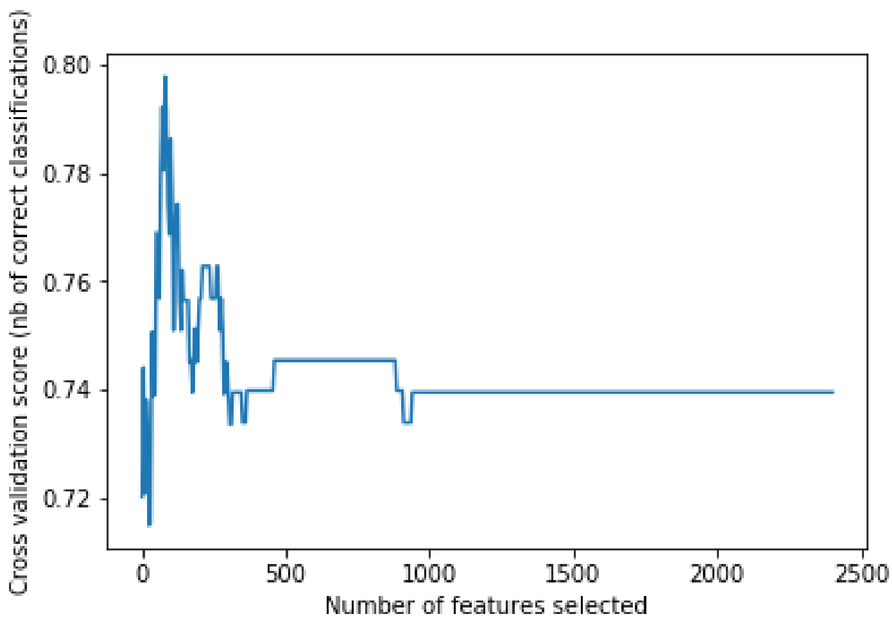

Figure 4.7: Graphic provided by SVM-RFE detailing its process.

From the observation of the first times the SVM-RFE was ran, two optional entry parameters were changed from the default values: the minimum value of features was capped at 30 and the step (number of features eliminated in each iteration) was set to 5 due to computational resources.

4.4 Conclusion

The present chapter described the main phase of this work. It begins with a more theoretical approach about the mathematical and statistical concepts that sustain the binary classification to be performed. Two different types of classification pipelines will take place: a ranking and selection of the top relevant features, and a recursive elimination of the less important feature. At last, the performance metrics are to be evaluated.

CHAPTER 5

Results

This chapter shows the application of the previous presented classification pipelines to cohort of data from 28 participants. The performance of the classification models will be displayed and compared against each other using both tables and graphics.

The selected features will also be studied in order to find out which ones appear more often and, therefore, may be the most important for the present case study. They will be analyzed by their time placement, scalp localization and their type, whether they are frequency or time features.

5.1 Experimental Set Up Strategy

5.1.1 Classifiers

As described in 3.1, in each trial, a participant was shown the face of either a target or a distractor and should provide an answer by trying to identify which of the two labels the picture matches. From all the duration of the trial, EEG signals were acquired and, after the feature extraction phase, a feature vector for each trial was created.

Each frequency vector belongs to one of four different classes: target and right answer, target and wrong answer, distractor and right answer, and distractor and wrong answer, which will be the labels for the supervised machine learning model. However, the number of wrong answers provided by the participants were in very small number when comparing with the right answers. In order to take the best advantage out of the available dataset, the main problem to attack will be the classification of target or distractor from the universe of right answers which is a binary classification problem. As in every binary problem there are only two answers, a 50% result or less is considered bad. In this work, it will be considered a metric value around 70% acceptable and a 90% result a very good result. Later on, it will also be tested if a machine learning model trained by with the right answers can correctly predict the wrong answers which is, in other words, identifying target and distractor where the participant has failed to do so.

Furthermore there should be approximately five distractor trials for each target trial. This means that the dataset at hands (see 3.1) is very unbalanced which usually affects the behavior of the machine learning model. Studies like [59] [60] have as their aim to solve the problem on how to approach unbalanced dataset for machine learning problems. Two different approaches were taken because of this factor:

- The dataset was distributed 50-50 and therefore balanced. In order to do that, for each participant, the number of distractors was reduced in order to match the number of targets. The priority here was given to the first trials based on the fact that, when a subject performs a task, over time the detection rate of targets tends to decrease [61].
- The model was fed with all the dataset. The SVM can be set to automatically adjust weights inversely proportional to the classes frequencies in the input data and with this deal with unbalanced datasets [62].

The classifiers will receive as entry a feature vector from one of two types: a frequency feature vector or a time feature vector. The first provides information about the trial's frequency band. The last one provides information related to the trial ERP components. The main focus will be to train and test the models with intra-participant data. Later on, inter-participants models will be tested as well.

5.2 Classification Pipeline Results

This section presents a series of results tables which aim to summarize the obtained results for intra-participant experiments. On each table, either an average accuracy/balanced accuracy is present. This value was obtain by performing the next steps:

- A cross-validation was performed in each participant which resulted in an average accuracy/balanced accuracy value each.
- An average across all the participants' accuracy/balanced accuracy values was performed and registered on the table.

Then was registered a standard deviation from the group of accuracy/balanced accuracy values as well. Finally, the highest and lowest values recorded were also registered as they mark the maximum and minimum performance obtained.

At the end of 5.2.1, further evaluations are performed. Note that in both of them, no cross-validation was performed because the nature of the evaluations inherently divides the data into trainning and testing groups.

5.2.1 Frequency-domain

The table below shows a brief summary of all the methods used for dealing with the frequency-domain balanced datasets, the respective average accuracy, which was the selected metric for comparison, and some additional measures. After some initial testing and observations, the ideal number of features to acquire from the ANOVA ranking ended up being around 100. As for the SVM-RFE, the number of selected features differed greatly with the average being 173.

It should be reminded that the classes used for this classification process were target+right and distractor+right.

	Average Accuracy	Accuracies' Standard Deviation	Maximum Accuracy	Minimum Accuracy
ANOVA w/ RBF kernel	0.853	0.073	0.983	0.717
ANOVA w/ linear kernel	0.819	0.097	1.000	0.633
SVM-RFE	0.735	0.115	0.943	0.515

Table 5.1: Classification of the frequency data on balanced datasets. Result values are displayed from 0 to 1.

The results when dealing with the unbalanced datasets weren't as promising as the ones acquired when working with the balanced datasets. The table 5.2 shows a brief summary of the results from the two techniques used for feature selection applied to the unbalanced dataset.

	Average Balanced Accuracy	Balanced Accuracies' Standard Deviation	Maximum Balanced Accuracy	Minimum Balanced Accuracy
ANOVA w/ RBF kernel	0.349	0.162	0.841	0.029
SVM-RFE	0.610	0.083	0.918	0.492

Table 5.2: Classification of the frequency data on unbalanced datasets. Result values are displayed from 0 to 1.

In contrast to what happened with the balanced datasets, the ANOVA technique with the 100 top features provided very bad results. This is no big surprise considering that the ANOVA method has proven to be more robust when dealing with small variations in data. As mentioned in 4.3.1, the ANOVA assumes that the data follows a normal distribution which is not the case for unbalanced data.

The SVM-RFE has proven to be more reliable in this situation. It can't be said that the results were good but it could indicate a proper starting point for future studies. The average number of features selected per model was 190.

5.2.1.1 Further evaluations: stimulus incorrectly identified

Taking into account the very promising results acquired in the frequency domain when working with balanced datasets, other tests were performed based on the most successful feature selection technique, the ANOVA.

Firstly, using as training data the trials where the participant answered correctly, the model developed was tested with the wrong decisions made by that same participant. In other words, to try and see if it is possible for the presented machine learning models to actually detect the presence of a target or a distractor even if the participants themselves did not. Unfortunately, the results weren't as good as the previous ones.

For testing, it was selected only participants with at least 20 trials on the less predominant class. The detailed results are shown in table 5.3. The average accuracy was set on 51% which is a poor result.

Number of participants	Average Accuracy	Accuracies' Standard Deviation	Maximum Accuracy	Minimum Accuracy
17	0.513	0.074	0.658	0.417

Table 5.3: Classification of the participants mistakes using the hits to train the model. Result values are displayed from 0 to 1.

5.2.1.2 Further evaluations: inter-participant

Secondly, twenty-eight SVM models with the same parameters as the specified in 4.3.1 were trained each one with the data from one of the twenty-eight eligible participants and then tested with the datasets from all other twenty-seven participants. The poor results displayed in the figure 5.1 proved right the conclusion supported by [4] that no conclusion could be drawn from all the participants all together and that an BCI is design and adapted for a specific participant. The graphic shows the mean accuracy and its respective standard deviation taken from each model's evaluation procedure.

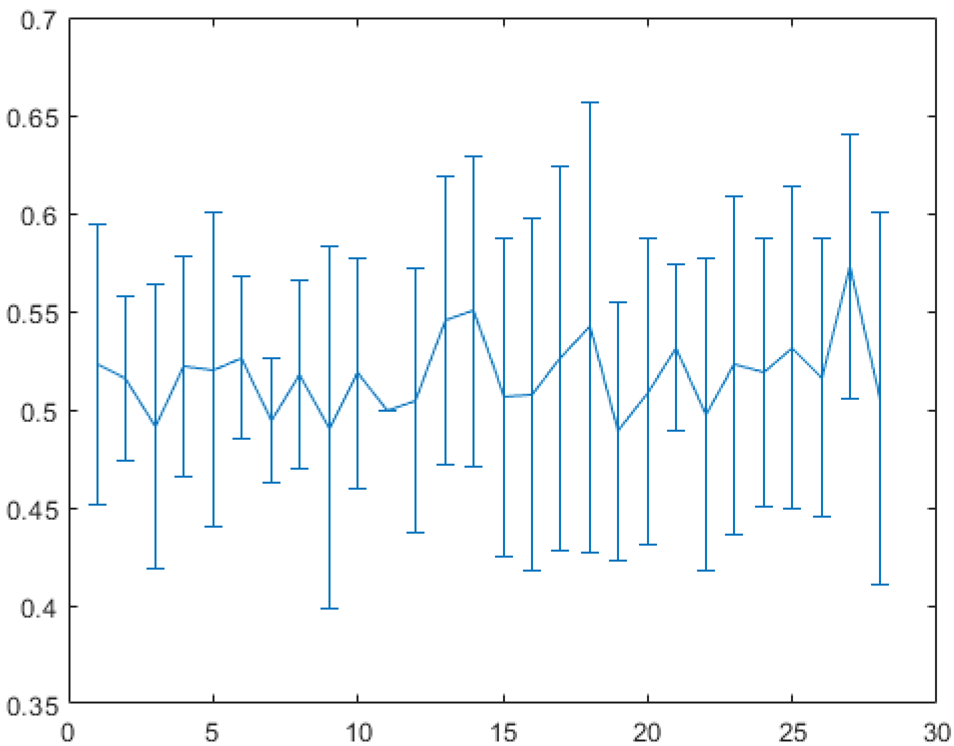

Figure 5.1: An errorbar [63] plot of 28 SVM models. On the xx axis are the number of the participant's data by which the model was trained. It then was tested on the other 27 datasets. The mean accuracy is represented on the yy axis.

5.2.2 Time-domain

The table 5.4 below compares the two methods used for selecting the more relevant time features for the balanced dataset. In this case, the ideal number of features to acquire from the ANOVA ranking ended up being around 50. Again, the number of features selected by the SVM-RFE differed greatly and this time the average was of 78 features. The average accuracy is lower with the time features when comparing it with the frequency features' average accuracy but in both cases the ANOVA method has proven to be more efficient than the SVM-RFE.

	Average Accuracy	Accuracies' Standard Deviation	Maximum Accuracy	Minimum Accuracy
ANOVA w/ RBF kernel	0.713	0.051	0.801	0.591
ANOVA w/ linear kernel	0.657	0.066	0.783	0.495
SVE-RFE	0.618	0.076	0.750	0.453

Table 5.4: Classification of the time data on balanced datasets. Result values are displayed from 0 to 1.

Table 5.5 details the performance of the same two methods of feature selections used earlier on and, just like with the frequency features, the ANOVA, with 50 features this time, had a very bad performance. However, with the time features, the SVM-RFE not only proved to be a lot more efficient than the pure statistical method but also achieved the best overall result so far, even better than it had achieved with the balanced dataset and with an average number of 31 selected features.

	Average Balanced Accuracy	Balanced Accuracies' Standard Deviation	Maximum Balanced Accuracy	Minimum Balanced Accuracy
ANOVA w/ RBF kernel	0.250	0.097	0.478	0.033
SVM-RFE	0.880	0.025	0.930	0.833

Table 5.5: Classification of the time data on unbalanced datasets. Result values are displayed from 0 to 1.

5.2.3 Frequency-Time Fusion

Following the knowledge acquired so far, the following points can be inferred:

- Overall, the best combined results between frequency and time features were achieved by working with a balanced dataset.
- On a balanced dataset, for the frequency features, the best result was achieved by selecting the 100 most important ones using the ANOVA method.
- On a balanced dataset, for the time features, the best result was achieved by selecting the 50 most important ones using the ANOVA method.

Now, a question can be asked: What would be the performance of a model if it was built by using both the 100 frequency features and the 50 time features?

The final phase of the present work teams up the frequency and time features for the most successful case so far, the ANOVA method on balanced datasets and it aims to discover if the results would be better than each one of them separately. The implementation steps of this model were similar to the ones represented in the diagram of figure 4.5 and a RBF kernel was used in the SVM model. Table 5.6 displays the results achieved from this fusion of data.

	Average Accuracy	Accuracies' Standard Deviation	Maximum Accuracy	Minimum Accuracy
ANOVA (freq+time)	0.864	0.073	1.000	0.700

Table 5.6: Classification of the frequency and time data combined on balanced datasets. Result values are displayed from 0 to 1.

It can be noted a slight improvement in the average accuracy when comparing with the frequency features alone. One of the participants actually achieved the top result of one hundred percent accuracy.

5.3 Feature Relevance

For this section, only the features selected by the ANOVA method across balanced datasets will be studied. The ANOVA method is preferred over the SVM-RFE because the number of features selected is always the same. In the case of SVM-RFE, the best scenario can be achieved by using a total number of features which can vary a lot across all participants, making it difficult to compare them and finding the most common features. As explained before, the ANOVA doesn't work well with unbalanced datasets which leave the balanced datasets to work with.

5.3.0.1 Electrode Pooling

When analysing the information of the four scalp lobes, channels were aggregated as follows, with the channel names being the used criteria:

- Frontal lobe: Fp1, Fp2, Fpz, F1, F3, F4, F8, Fz, FC5, FC1, FC2 and FC6.
- Parietal lobe: Cz, C3, C4, CP1, CP2, CP5, CP6, Pz, P3, P4, P7 and P8.
- Temporal lobe: T7 and T8.
- Occipital lobe: POz, O1, O2 and Oz.

5.3.1 Frequency-domain

The pie chart displayed in figure 5.2 shows the average predominance of the three frequency bands (theta, alpha and beta) selected by the ANOVA method. Each slice is the average of that wave throughout all participants.

The alpha band was the less predominant one of all three, which goes accordingly to the information provided in [11] that was described in 2.2.1.1. The theta band was expected to dominate the graph due to its relation to memory, specially with episodic memory as shown in studies [64] and [65], but that place was claimed by the beta band, specially in the temporal

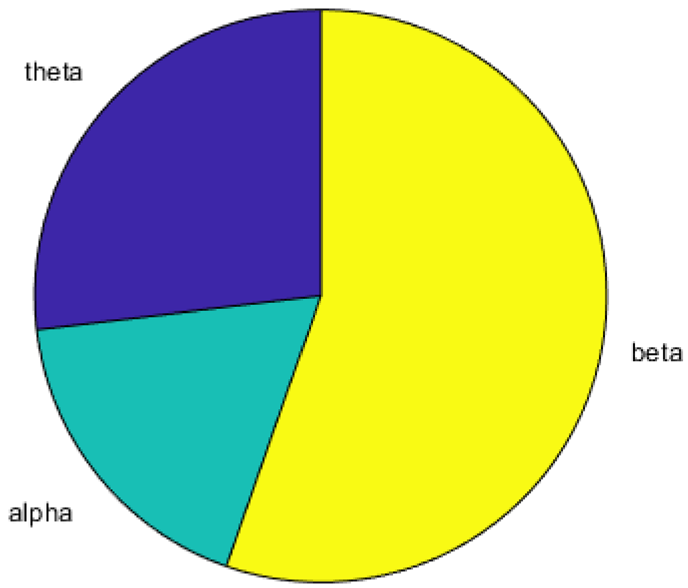

Figure 5.2: Average predominance of theta, alpha and beta bands in the one hundred features selected by the ANOVA method for all participants.

channels (T7 and T8) which can be seen in figure 5.3. This figure represents the 148 most selected features with each one appearing in at least 4 participants.

Although not having a clean sheet like in the temporal lobe, the beta frequency is also very commonly found in the frontal lobe.

The predominance of channels in the temporal zone wasn't surprising as [64] and [65] link them to episodic memory.

One can also point out that most of the selected features were within the first second post-stimulus. Despite the veracity of this claim, it is not very salient. What is also claimable is that less features are chosen as time goes by. Table 5.7 shows the average from all participants of the percentage of features chosen in each time period. Note that the time windows before stimulus is only half a second.

before stimulus	1st second	2nd second	3rd second
0.096	0.378	0.291	0.235

Table 5.7: Average percentages from all participants of the selected features in each time windows. Percentages values are displayed from 0 to 1.

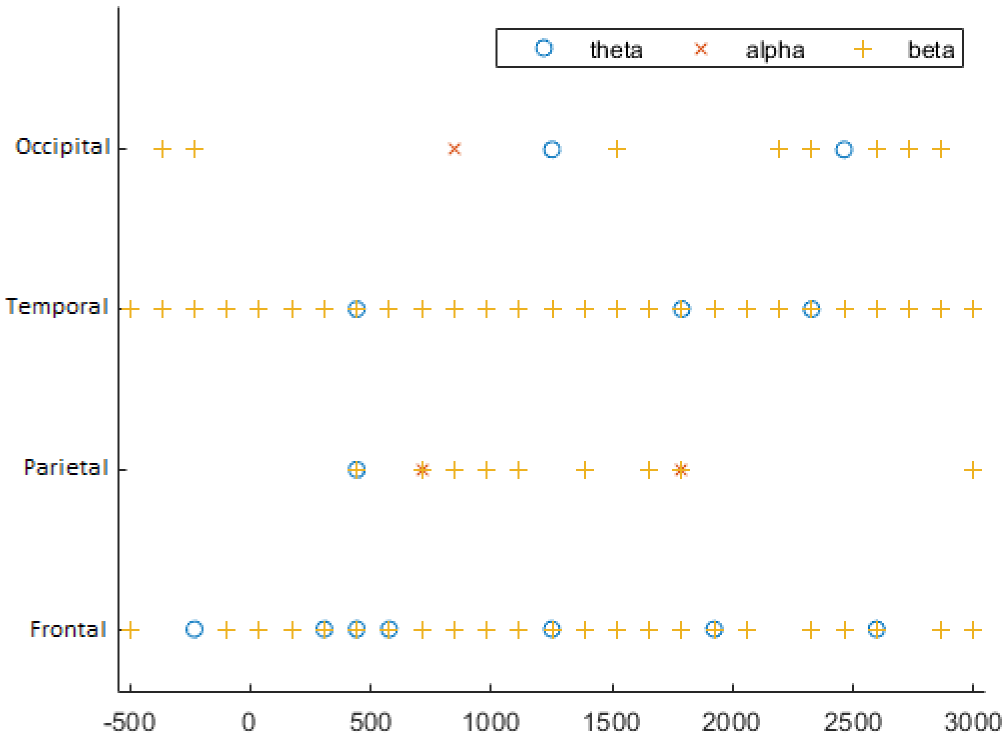

Figure 5.3: Most common frequency features selected by the ANOVA method throughout all the models. The horizontal axis represents the epoch time in milliseconds being 0 the instant of the stimulus.

5.3.1.1 Spatial and Temporal Feature Localization

For each participant, three binary matrices were extracted, for theta, alpha and beta, with the columns being the time instants and the lines being the channels selected by the feature selection algorithm. This allowed to go back on MATLAB and, with the help of EEGLAB, make a continuous stream of images that better show which channel is being selected as an important feature in each different time interval. An example of this stream can be seen in figure 5.4 where it's represented all the frequency features selected in the beta band throughout all the 27 time intervals.

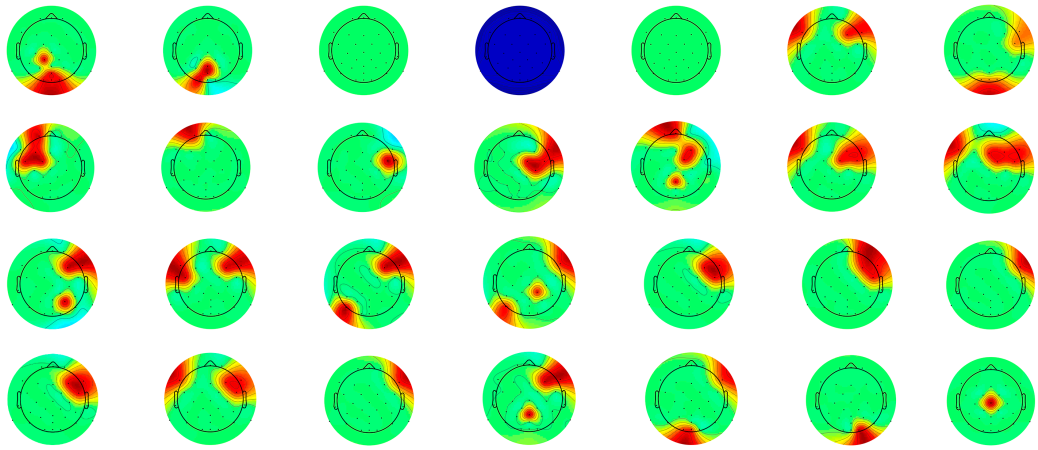

Figure 5.4: All frequency features in the beta band from a participant in every time interval, ordered from left to right and from top to bottom and the blue one represents the occurrence of the stimulus. The fourth(blue) and fifth topoplots occurred at the same time instant.

5.3.1.2 Further evaluations: scalp temporal beta

After analysing the feature relevance, the scalp temporal features in the beta frequency band were noticed to be present in great number across all the participants. So curiosity related to how would a model with just the two temporal channels in the beta band led to this final experiment in the frequency-domain with balanced data. The results aren't nearly as good as the one performed with the best one hundred features selected by the ANOVA technique. However, if taken into account the extreme simplicity of the procedure, just two electrodes and their energy points between [13, 30)Hz which produce a total of only 58 features, there are some impressive accuracy values in display in table 5.8. Despite the average accuracy value, the range and even the maximum accuracy value show promising results for possible future works. The SVM model was built with a RBF kernel.

	Average Accuracy	Accuracies' Standard Deviation	Maximum Accuracy	Minimum Accuracy
Beta T7 and T8	0.689	0.120	0.960	0.509

Table 5.8: Classification of the frequency data on balanced datasets using only the beta frequency band on the temporal channels. Result values are displayed from 0 to 1.

5.3.2 Time-domain

Figure 5.5 shows the 36 most predominant features that appear in at least 12 participants when using ANOVA for feature selection. Unlike what happened before with the beta band in the temporal zone, no particular feature stands out from the rest. The amplitude appears to be generally more important than the latency and the P300 is present in every region. The temporal lobe has a notable lack of time features. This fact actually makes sense if taken into account that beta frequencies were the predominant ones in this lobe. Let's remember that, in 3.4, during the smoothing process of the time data, the beta band is actually eliminated.

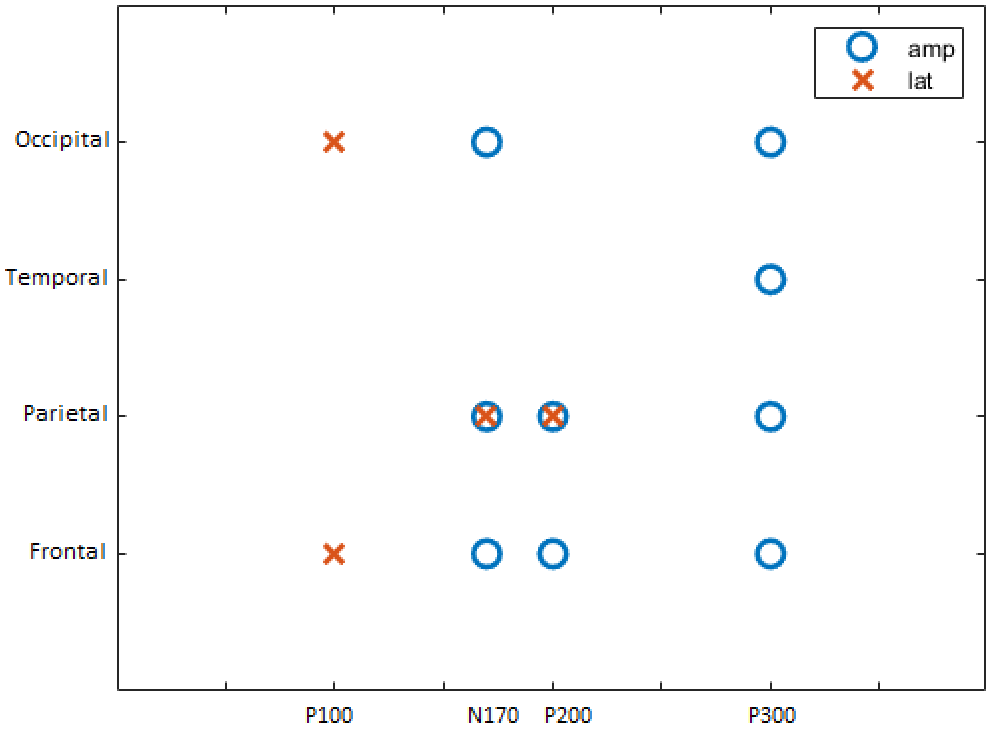

Figure 5.5: Most common time features selected by the ANOVA method throughout all the models. On the horizontal axis are represented the ERP components of interest.

5.4 Conclusion

This chapter presented the application of the classification pipelines described in 4.3 to the feature vectors described in 3.5. The nature of the predominant features, both the spatial and temporal characteristics, were studied but only the ones provided by the ANOVA method since the number of selected features by the SVM-RFE was highly different among all the participants.

CHAPTER 6

Conclusion

In this work, a solution for a BCI that deals with a criminal scenario was proposed. The work was performed with data extracted from 28 participants which witness criminal videos showing a crime being committed and then had to identify the culprit in several sequences of pictures.

Firstly, the EEG signals were grouped and given a respective class from the following:

- target and right answer;
- target and wrong answer;
- distractor and right answer;
- distractor and wrong answer.

Next, information from both frequency and time domain was extracted from the EEG signals. Frequency domain data consisted in information from theta, alpha and beta band during the entire epoch and time domain information consisted in the power and latency values of the following ERP components: P100, N170, P200 and P300.

The gathered data was injected into standard SVM models using two types of feature selection techniques: a feature ranking with the ANOVA method and a feature recursive elimination with the SVM-RFE. Tests were performed using both balanced and unbalanced datasets, with features from the frequency domain only, with features from the time domain only and, finally, with features from both time and frequency domain.

From the observation of the results, the following conclusions are drawn:

- The frequency features have shown very impressive results, with some dominance from the beta band. This may came with a bit of surprise since studies like [64] and [65] take the theta band as the most important frequency band in the subject of episodic memory.
- The results acquired from only the two temporal channels in just the beta frequency band were at least interesting. Of course, the overall accuracy cannot be compared with the values obtained using all frequency data, however, for such a simple model, the results should be taken into account for future studies.
- Despite the good results provided by the frequency features in classifying as target or distractor samples which the participant had given a correct answer, when trying to use

this same data to train a model and use it to classify the participant's wrong answers, the results weren't as promising. In the end, the model ended up picking conscious thinking of a person. Perhaps by looking at this problem from a different perspective or with different kinds of features it could be possible to draw some conclusion but not with the methods used in this work.

- Trying to perform inter-participant studies doesn't provide results at all. This is not a surprise since it's common for a BCI to have a period of testing and adapting to a given person [4].
- It is also with no surprise that the results provided by balanced datasets are better than the ones from unbalanced datasets despite the SVM allowing to change the weight from the different classes. There are in fact studies [59] [60] which only aim is to determine what is the best way to handle unbalance dataset and, in most scenarios, the solution ends up being turning the unbalance dataset into a balanced one, either by sample reduction or by forging new samples based on the existing ones. This was verified when dealing with the frequency features but was definitely not the case for the time features. The balanced accuracy obtained by the SVM-RFE for unbalanced datasets even surpassed the accuracy obtained by the SVM-RFE itself when using balanced datasets.

6.1 Future Work

In chapter 1, it was referred that around seventy one percent of the innocent convicted criminals had been condemned by eye witnesses. However, the participants' misclassifications represents about ten percent of the entire database. More experiments should be conducted in order to find out if this number is actually the general case. Also, since the dataset is so unbalance, one can focus its works on dealing with just the unbalanced data.

The present work focused frequency features with the theta, alpha and beta frequency bands as well as four different types of ERP components. Every study was based on single trials but other approaches may be taken instead like, for example, studying the average of two or three signals and treat them like single trials. This is proved [66] this to be useful for study purposes, however it cannot be used for correcting classification in real time since it can't look at just a single unique sample.

Perhaps the most notorious conclusion of the work is the constant appearance of the beta band in the temporal channels which is impossible to miss. It could be the starting point of a future work, focused on studying the importance of these frequency features for the present topic.

As mentioned before, a different approach is going to have to be taken if one is trying to classify correctly the misclassifications from a participant. With the analysis performed, it was only possible to clear detect the intention of the participant. The topic of unconscious recognition of a face of a culprit would came as a huge advantage in any justice system and it surely deserves more investigation.

References

[1] *Eyewitness identification reform.* [Online]. Available: `https://www.innocenceproject.org/eyewitness-identification-reform`, (accessed: 02.12.2019).

[2] A. Bablani, D. R. Edla, D. Tripathi, and R. Cheruku, "Survey on Brain-Computer Interface", *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–32, 2019, ISSN: 03600300. DOI: `10.1145/3297713`.

[3] B.-h. Gabriel, "PEDRO JOÃO PSICOFISIOLOGIA DO TESTEMUNHO OCULAR PSYCHOPHYSIOLOGY OF EYEWITNESS TESTIMONY Universidade de Aveiro Departamento de Educação e Psicologia PEDRO JOÃO BEM-HAJA GABRIEL PSICOFISIOLOGIA DO TESTEMUNHO OCULAR PSYCHOPHYSIOLOGY OF EYEWITNESS", PhD thesis, 2018.

[4] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update", *Journal of Neural Engineering*, vol. 15, no. 3, 2018, ISSN: 17412552. DOI: `10.1088/1741-2552/aab2f2`.

[5] S. N. Abdulkader, A. Atia, and M. S. M. Mostafa, "Brain computer interfacing: Applications and challenges", *Egyptian Informatics Journal*, vol. 16, no. 2, pp. 213–230, 2015, ISSN: 11108665. DOI: `10.1016/j.eij.2015.06.002`. [Online]. Available: `http://dx.doi.org/10.1016/j.eij.2015.06.002`.

[6] A. Segev, D. Curtis, S. Jung, and S. Chae, "Invisible brain: Knowledge in research works and neuron activity", *PLoS ONE*, vol. 11, no. 7, pp. 1–21, 2016, ISSN: 19326203. DOI: `10.1371/journal.pone.0158590`.

[7] M. Sachs, "Anatomy of the brain", *Soins; la revue de reference infirmiere*, vol. 27, no. 2, pp. 3–8, 1982, ISSN: 0038-0814.

[8] L. M. Ross and E. D. Lamperti, *Head and Neuroanatomy: Thieme Atlas of Anatomy*, 1. 2008, vol. 29, pp. 207–208, ISBN: 9781604062908. DOI: `10.3174/ajnr.a0733`.

[9] *Lobes of the brain.* [Online]. Available: `https://courses.lumenlearning.com/waymaker-psychology/chapter/reading-parts-of-the-brain/`, (accessed: 01.06.2020).

[10] P. Estupinya, "Como funciona o cérebro", *National Geographic*, 2019.

[11] IMotions, *Electroencephalography - The Complete Pocket Guide*, 2019.

[12] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10-20 system electrode placement", *Electroencephalography and Clinical Neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987, ISSN: 00134694. DOI: `10.1016/0013-4694(87)90206-9`.

[13] *10-20 system of electrode placement.* [Online]. Available: `https://faculty.washington.edu/chudler/1020.html`, (accessed: 16.06.2020).

[14] J. J. Gibson, "The concept of the stimulus in psychology.", *American Psychologist*, vol. 15, no. 11, pp. 694–703, 1960, ISSN: 0003-066X. DOI: `10.1037/h0047037`.

[15] M. Ramzan and S. Dawn, "A survey of brainwaves using electroencephalography(EEG) to develop Robust Brain-Computer Interfaces(BCIs): Processing techniques and Algorithms", *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, pp. 642–647, 2019. DOI: `10.1109/CONFLUENCE.2019.8776890`.

[16] *Theta brain state.* [Online]. Available: `https://www.thebigart.com/theta-brain-state`, (accessed: 13.06.2020).

[17] S. J. Luck, “An Introduction to Event-Related Potentials and Their Neural Origins”, *An introduction to the event-related potential technique*, 2005, ISSN: 14657309. DOI: 10.1007/s10409-008-0217-3. arXiv: arXiv:1011.1669v3.

[18] S. Sanei and J. A. Chambers, *EEG Signal Processing*. 2013, ISBN: 9780470025819. DOI: 10.1002/9780470511923.

[19] E. S. Kappenman and S. J. Luck, *The Oxford Handbook of Event-Related Potential Components*. 2012, ISBN: 9780199940356. DOI: 10.1093/oxfordhb/9780195374148.001.0001.

[20] *Researchers develop a fast, noninvasive brain-computer interface*. [Online]. Available: https://medicalxpress.com/news/2015-10-fast-noninvasive-brain-computer-interface.html, (accessed: 01.06.2020).

[21] N. Brodu, F. Lotte, and A. Lécuyer, “Comparative study of band-power extraction techniques for Motor Imagery classification”, *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - CCMB 2011: 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain*, pp. 95–100, 2011. DOI: 10.1109/CCMB.2011.5952105.

[22] *Visualize machine learning data in python with pandas*. [Online]. Available: https://machinelearningmastery.com/visualize-machine-learning-data-python-pandas/, (accessed: 11.06.2020).

[23] C. A. G. Marcelo, Z. R. B. Pasquin, A. D. T. Pichay, M. L. D. Tan, M. F. K. N. Simon, S. V. Prado, C. V. Buenaventura, and M. S. Nicasio, “Characterization and comparison of brain wave signals during deception”, in *HNICEM 2017 - 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, 2018, ISBN: 9781538609101. DOI: 10.1109/HNICEM.2017.8269508.

[24] A. Hassan, J. Tao, G. Li, M. Jiang, L. Aii, J. Zhihui, L. Zongfang, and C. Qibing, “Effects of Walking in Bamboo Forest and City Environments on Brainwave Activity in Young Adults”, *Evidence-Based Complementary and Alternative Medicine*, 2018, ISSN: 1741-427X. DOI: 10.1155/2018/9653857.

[25] *Understanding analysis of variance (anova) and the f-test*. [Online]. Available: https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test, (accessed: 11.06.2020).

[26] *Supervised and unsupervised machine learning algorithms*. [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/, (accessed: 29.01.2020).

[27] *Machine learning classifiers*. [Online]. Available: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623, (accessed: 12.02.2020).

[28] *Random forest algorithm for machine learning*. [Online]. Available: https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb, (accessed: 12.02.2020).

[29] *Introduction to machine learning model evaluation*. [Online]. Available: https://heartbeat.fritz.ai/introduction-to-machine-learning-model-evaluation-fa859e1b2d7f, (accessed: 12.02.2020).

[30] *A gentle introduction to k-fold cross-validation*. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/, (accessed: 13.02.2020).

[31] P. A. A. Resende and A. C. Drummond, “A survey of random forest based methods for intrusion detection systems”, *ACM Computing Surveys*, vol. 51, no. 3, 2018, ISSN: 15577341. DOI: 10.1145/3178582.

[32] *Preferable tools for machine learning - python - matlab - r*. [Online]. Available: https://www.codementor.io/@innat_2k14/preferable-tools-for-machine-learning-python-matlab-r-jfozzpphz, (accessed: 29.01.2020).

[33] *What is eeglab?* [Online]. Available: https://sccn.ucsd.edu/eeglab/index.php, (accessed: 25.01.2020).

[34] *Feature extraction*. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/feature-extraction, (accessed: 09.12.2020).

[35] *Zero-phase filtering*. [Online]. Available: https://flylib.com/books/en/2.729.1/zero_phase_filtering.html, (accessed: 25.05.2020).

[36] E. E. Smith, S. J. Reznik, J. L. Stewart, and J. J. Allen, "Assessing and conceptualizing frontal EEG asymmetry: An updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry", *International Journal of Psychophysiology*, vol. 111, pp. 98–114, 2017, ISSN: 18727697. DOI: 10.1016/j.ijpsycho.2016.11.005. [Online]. Available: http://dx.doi.org/10.1016/j.ijpsycho.2016.11.005.

[37] M. V. Martin, V. Cho, and G. Aversano, "Detection of subconscious face recognition using consumer-grade brain-computer interfaces", *ACM Transactions on Applied Perception*, vol. 14, no. 1, 2016, ISSN: 15443965. DOI: 10.1145/2955097.

[38] M. Hoshiyama, R. Kakigi, S. Watanabe, K. Miki, and Y. Takeshima, "Brain responses for the subconscious recognition of faces", *Neuroscience Research*, vol. 46, no. 4, pp. 435–442, 2003, ISSN: 01680102. DOI: 10.1016/S0168-0102(03)00121-4.

[39] C. Nan, G. Wang, H. Wang, X. Wang, Z. Liu, L. Xiao, H. Bai, and S. Wu, "The P300 component decreases in a bimodal oddball task in individuals with depression: An event-related potentials study", *Clinical Neurophysiology*, 2018, ISSN: 18728952. DOI: 10.1016/j.clinph.2018.09.012.

[40] C. A. Frantzidis, C. Bratsas, M. A. Klados, E. Konstantinidis, C. D. Lithari, A. B. Vivas, C. L. Papadelis, E. Kaldoudi, C. Pappas, and P. D. Bamidis, "On the classification of emotional biosignals evoked while viewing affective pictures: An integrated data-mining-based approach for healthcare applications", *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 309–318, 2010, ISSN: 10897771. DOI: 10.1109/TITB.2009.2038481.

[41] *Feature vector*. [Online]. Available: https://brilliant.org/wiki/feature-vector/, (accessed: 01.10.2020).

[42] T. Evgeniou and M. Pontil, "Workshop on Support Vector Machines : Theory and Applications", *Machine Learning and Its Applications: Advanced Lectures*, no. January 2001, pp. 249–257, 2001. DOI: 10.1007/3-540-44673-7.

[43] *A friendly introduction to support vector machines(svm)*. [Online]. Available: https://towardsdatascience.com/a-friendly-introduction-to-support-vector-machines-svm-925b68c5a079, (accessed: 25.06.2020).

[44] *Support vector machines*. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html, (accessed: 21.08.2020).

[45] *Sklearn.feature$_s$election.rfe*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html, (accessed: 20.07.2020).

[46] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining (New International Editon)*. 2013, ISBN: 0-321-32136-7. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

[47] *Sklearn.feature$_s$election.rfecv*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFECV.html, (accessed: 20.07.2020).

[48] Z. Li, W. Xie, and T. Liu, "Efficient feature selection and classification for microarray data", *PLoS ONE*, vol. 13, no. 8, pp. 1–21, 2018, ISSN: 19326203. DOI: 10.1371/journal.pone.0202167.

[49] A. R. Hidalgo-Muñoz, M. M. López, I. M. Santos, A. T. Pereira, M. Vázquez-Marrufo, A. Galvao-Carmona, and A. M. Tomé, "Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing", *Expert Systems with Applications*, vol. 40, no. 6, pp. 2102–2108, 2013, ISSN: 09574174. DOI: 10.1016/j.eswa.2012.10.013.

[50] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks", *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009, ISSN: 03064573. DOI: 10.1016/j.ipm.2009.03.002. [Online]. Available: http://dx.doi.org/10.1016/j.ipm.2009.03.002.

[51] *Sklearn.metrics.balanced$_a$ccuracy$_s$core*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html, (accessed: 09.07.2020).

[52] X. Wu, V. Kumar, Q. J. Ross, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining", *Knowledge and Information Systems*, 2008, ISSN: 02191377. DOI: 10.1007/s10115-007-0114-2.

[53] *Scikit-learn*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html, (accessed: 13.07.2020).

[54] *Sklearn.metrics.accuracy$_s$core*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html, (accessed: 13.07.2020).

[55] *Analyzing differences between groups*. [Online]. Available: https://writing.colostate.edu/guides/page.cfm?pageid=1398&guideid=67, (accessed: 01.09.2020).

[56] *Sklearn.feature$_s$election.selectkbest*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html, (accessed: 01.09.2020).

[57] *Normal distribution*. [Online]. Available: https://www.mathsisfun.com/data/standard-normal-distribution.html, (accessed: 17.08.2020).

[58] *Sklearn.preprocessing.standardscaler*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html, (accessed: 19.08.2020).

[59] H. Núñez, L. Gonzalez-Abril, and C. Angulo, "Improving SVM Classification on Imbalanced Datasets by Introducing a New Bias", *Journal of Classification*, vol. 34, no. 3, pp. 427–443, 2017, ISSN: 14321343. DOI: 10.1007/s00357-017-9242-x.

[60] N. J. Rehan Akbani, Stephen Kwek, "Applying Support Vector Machines to Imbalanced", no. July, 2015. DOI: 10.1007/978-3-540-30115-8.

[61] *Alertness*. [Online]. Available: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/alertness, (accessed: 20.07.2020).

[62] *Sklearn.svm.svc*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html, (accessed: 22.07.2020).

[63] *Errorbar*. [Online]. Available: https://www.mathworks.com/help/matlab/ref/errorbar.html, (accessed: 09.09.2020).

[64] S. Hanslmayr, B. P. Staresina, and H. Bowman, *Oscillations and Episodic Memory: Addressing the Synchronization/Desynchronization Conundrum*, 2016. DOI: 10.1016/j.tins.2015.11.004.

[65] S. C. Berens and A. J. Horner, *Theta Rhythm: Temporal Glue for Episodic Memory*, 2017. DOI: 10.1016/j.cub.2017.08.048.

[66] A. M. S. Steven J. Luck Gi-Yeul Bae, *Overview of erp decoding*.