MARCO ANTONIO PINTO ORELLANA

# SEIZURE DETECTION IN ELECTROENCEPHALOGRAMS USING DATA MINING AND SIGNAL PROCESSING

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa**

T

P659s
2017

Pinto Orellana, Marco Antonio, 1991-
Seizure detection in electroencephalograms using data mining and signal processing / Marco Antonio Pinto Orellana. – Viçosa, MG, 2017.
xii, 69f. : il. (algumas color.) ; 29 cm.

Orientador: Fábio Ribeiro Cerqueira.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.59-69.

1. Mineração de dados (Computação). 2. Epilepsia. 3. Eletroencefalografia de ritmo alfa. 4. Doenças mentais. I. Universidade Federal de Viçosa. Departamento de Informática. Programa de Pós-graduação em Ciência da Computação. II. Título.

CDD 22 ed. 006.312

MARCO ANTONIO PINTO ORELLANA

# SEIZURE DETECTION IN ELECTROENCEPHALOGRAMS USING DATA MINING AND SIGNAL PROCESSING

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 10 de março de 2017.

Ricardo dos Santos Ferreira

Marcos Henrique Fonseca Ribeiro

Fabio Ribeiro Cerqueira
Orientador

*A mis padres, Marco y María, y a mi hermanita, Cris, por su siempre incondicional presencia e indeleble ayuda en los momentos más arduos, difíciles y oscuros. Todo el documento y estudio aquí presente jamás podría haberse realizado sin su apoyo y bendición.*

*"We've learned from experience that the truth will come out."*
(Richard Feynman)

# Acknowledgments

First and foremost, I want to thank my family for all their support, for their patience, listening to me about my crazy ideas and thoughts, for their presence that helps me never feel alone, and for helping me to stay motivated and enthusiastic. Without their support, this document would never have been finished.

Thanks to my friends for reminding me that it is possible to have a family that does not share our blood.

I am also grateful to my professors, for opening my eyes to new and amazing knowledge areas. And, especially, to my advisor, for his support during the research.

To *Fundação de Amparo à Pesquisa de Minas Gerais* (FAPEMIG), and *Universidade Federal de Viçosa* (UFV), thanks for financially supporting this research work.

I would also like to thank someone who sacrificed his own health in order to help me in everything I did, even if that meant to be burnt several times in the process: my everlasting friend, my laptop.

And finally, last but not least, thanks to everyone at *Departamento de Informática*. Their support helped me to understand that "maldade" could have an unlimited number of meanings, and not all of them related to evil. It was great, most of the time, sharing the laboratory with all of you during last two years.

Thanks for all your encouragement!

# Contents

# List of Figures

# List of Tables

# Abstract

PINTO ORELLANA, Marco Antonio, M.Sc., Universidade Federal de Viçosa, March 2017. **Seizure detection in electroencephalograms using data mining and signal processing** Adviser: Fabio Ribeiro Cerqueira.

Epilepsy is one of the most common neurological diseases and is defined as the predisposition to suffer unprovoked seizures. The World Health Organization estimates that 50 million people are suffering this condition worldwide. Epilepsy diagnosis implies an expensive and long process based on the opinion of specialist personnel about electroencephalograms (EEGs) and video recordings. We have developed two methods for automatic seizure detection using EEG and data mining. The first system is a patient-specific method that consists of extracting spectro-temporal features of 23 EEG channels, applying a dimension reduction algorithm, recovering the envelope of the signal, and creating a model using a random forest classifier. Testing this system against a large dataset, we reached 97% of specificity and 99% of sensitivity. Thus, our first proposal showed to have a great potential for diagnosis support in clinical context. The other developed system is a non-patient specific method that consists of selecting the differential signal of two electrodes, applying an array of filter banks to that signal, extracting time series features, and creating a predictive model using a decision tree. The performance of this method was 95% of specificity, and 87% of sensitivity. Although the performance is lower than previous proposals, due to the design conditions and characteristics, our method allows an easier implementation with low hardware requirements. Both proposals presented here, using distinct approaches, demonstrate to be seizure prediction alternatives with very satisfactory performances under different circumstances and requirements.

# Resumo

PINTO ORELLANA, Marco Antonio, M.Sc., Universidade Federal de Viçosa, março de 2017. **Detecção de convulsões em eletroencefalogramas usando mineração de dados e processamento de sinais** Orientador: Fabio Ribeiro Cerqueira.

A epilepsia é uma das doenças neurológicas mais comuns definida como a predisposição a sofrer convulsões não provocadas. A Organização Mundial da Saúde estima que 50 milhões de pessoas estão sofrendo esta condição no mundo inteiro. O diagnóstico de epilepsia implica em um processo caro e longo baseado na opinião de especialistas com base em eletroencefalogramas (EEGs) e gravações de vídeo. Neste trabalho, foram desenvolvidos dois métodos para a predição automática de convulsões usando EEG e mineração de dados. O primeiro sistema desenvolvido é um método específico para cada paciente (*patient-specific*) que consiste em extrair características espectro-temporais de todos os canais de EEG, aplicar um algoritmo de redução de dimensão, recuperar o envelope do sinal e criar um modelo usando um classificador *random forest*. Testando este sistema com um grande banco de dados de epilepsia, atingimos 97% de especificidade e 99% de sensibilidade. Assim, a primeira proposta mostrou ter um grande potencial para colaborar com o diagnóstico em um contexto clínico. O segundo sistema desenvolvido é um método não específico do paciente (*non-patient specific*) que consiste em selecionar o sinal diferencial de dois eletrodos, aplicar um vetor de bancos de filtros para esse sinal, extrair atributos de séries temporais e criar um modelo preditivo usando uma árvore de decisão CART. O desempenho deste método foi de 95% de especificidade e 87% de sensibilidade. Estes valores não são tão altos quanto os de métodos propostos anteriormente. No entanto, a abordagem que propomos apresenta uma viabilidade muito maior para implementação em dispositivos que possam ser efetivamente utilizados por pacientes em larga escala. Isto porque somente dois elétrodos são utilizados e o modelo de predição é computacionalmente leve. Note-se que, ainda assim, o modelo

gerado apresenta um poder preditivo satisfatório e generaliza melhor que em trabalhos anteriores já que pode ser treinado com dados de um conjunto de pacientes e utilizado em pacientes distintos (*non-patient specific*). Ambas as propostas apresentadas aqui, utilizando abordagens distintas, demonstram ser alternativas de predição de convulsões com performances bastante satisfatórias sob diferentes circunstâncias e requisitos.

# Chapter 1

# Introduction

Epilepsy is a common and chronic neurological disease derived from consecutive occurrence of unprovoked seizures [Guo et al., 2010; Awan et al., 2015; Bhavaraju et al., 2006; Fatichah et al., 2014; Duo Chen et al., 2015; Dong and Liang, 2014; WHO, 2006]. This illness is associated with a severe reduction in the quality of life of patients. The risk of death arises as a consequence of the epilepsy symptoms and, therefore, patients require special care [Czarnecki and Gustafsson, 2015]. The World Health Organization (WHO) estimates that there are near to 200 million people who suffer epilepsy worldwide, although officially only 50 million are registered [WHO, 2006]. A proper diagnosis could help to apply the medication therapy that can help to control the most critical symptoms. Nevertheless, due to the risks of these types of medicines, their prescription is strongly limited to accurate diagnosis by specialists in the field [WHO, 2006; Orosco et al., 2016].

Common epilepsy diagnosis relies on two methods used together: electroencephalograms (EEGs) and video monitoring during long periods of time [Page et al., 2015; Orosco et al., 2016]. The long term of the recordings reduces the efficiency of the analysis, because it is directly proportional to the consumed time that the health specialists need to revise and evaluate the data [Tsiouris et al., 2015]. Consequently, several alternatives were proposed to reduce the processing time ranging from circuitry solutions [Bhavaraju et al., 2006], implemented in analog devices, to software methods that depend on artificial intelligence techniques [Czarnecki and Gustafsson, 2015; Orosco et al., 2013].

Electroencephalograms are recordings at the scalp of the variations in the electrical current. Such a variations are related to the neuronal activity, particularly to the electrochemical interactions of neuronal subsets. But, those interactions are only able to be detected when the electrical activity is synchronized and executed for large ensembles of neurons [Sierra-Marcos et al., 2015; Ahmad et al., 2014; Azevedo et al., 2009; Page et al., 2015]. This measurement technique is exploited in several fields, including seizure detection, due to its low cost, noninvasiveness, and medical safety despite its low resolution (a few microvolts) and high sensitivity to noise [Leuthardt et al., 2006].

Solutions, in this field of research, were proposed using numerous combinations of methods. The vast majority extracts spectro-temporal features of the signal and processes that information using neural networks, or support vector machines as classification algorithms [Das et al., 2016; Rahul Kumar Chaurasiya et al., 2015; Guo et al., 2010]. Until 2015, studies with large datasets of epileptic seizures, such as the Children Hospital of Boston (CHB-MIT) database [Goldberger et al., 2000a], were able to recognize correctly from 88% to 92% of the seizure episodes and between 86% and 95% of the seizure-free intervals [Ahmad et al., 2014; Das et al., 2015; Iqbal et al., 2015; Tsiouris et al., 2015; Gill et al., 2015]. All of them were based on the analysis of EEG channels in the whole brain. The present study continues that data extraction methodology looking for improving the performance benchmarks in first place, and later, reducing the number of electrodes needed to achieve good prediction scores.

## 1.1 Problem and its importance

### 1.1.1 Problem identification

In biomedical datasets, data format directly affects the diagnosis procedure by specialized medical personnel. Simple and proper representations reduce time analysis, and increase the accuracy of the evaluation of the data. Medical specialists are trained to read raw brainwaves and classify them. They determine if some oscillations are considered as a variation of a normal condition, or, on the opposite, they represent a disorder in the normal operation of the brain [Vogel, 1970]. Sometimes, this task needs additional EEG tests to check whether some patterns and variations appear with more emphasis on the observed data [Benbadis, 2006].

Bendabis et al. pointed that, in the case of epilepsy analysis, misdiagnosis is a relatively common situation. Nearly one-third of patients in epilepsy centers do not suffer epilepsy, but a kind of seizure derived from other illnesses [Benbadis and Lin, 2008]. This is one of the reasons why a proper epilepsy diagnosis must include the medical history, and EEG recordings along with video monitoring, to confirm the medical opinion [Page et al., 2015; Orosco et al., 2016].

In general, EEG processing is not a simple task due to three factors related with the acquisition and digitization process. First, the low amplitude of signals measured by the electrodes requires special devices (known as low-noise amplifiers) that do not introduce so much noise to the original signal [Gill et al., 2015]. Second, in the seizure context, small patient movements could interfere in the EEG measurement mechanism as another kind of noise [Bhagwat and Jain, 2013]. Last but not least, EEG properties like nonlinearity and nonstationarity make difficult to develop accurate prediction using the typical methods [Birjandtalab et al., 2016; Jovic et al., 2013].

The current study was developed to find new features of EEG signals to increase the effectiveness of epilepsy seizure prediction, reducing the processing requirements. For this purpose, our work was distributed in two main tasks accomplished with distinct methods.

### 1.1.2 Method I: Nonlinear features in dimension-reduced EEG

#### 1.1.2.1 Problem formulation

*Is it possible to detect seizures from a one-dimensional transformation of the electroencephalographic signals?*

#### 1.1.2.2 Justification

Epilepsy seizure recognition is a field of study addressed from several perspectives: signal domain changes, feature extraction, and data transformations. The information inside brainwaves is stored in nonlinear changes in the frequency spectrum signal. Due to this nature, different studies in the area typically transform the recorded data from a time domain into a frequency domain, or a spectro-temporal domain. To achieve this task, techniques such as Fourier transforms, Wavelet transforms, and short-time Fourier transforms are used [Das et al., 2016; Rahul Kumar Chaurasiya et al., 2015; Guo et al., 2010].

However, there is also another perspective to detect the signs of this disease: using of mathematical transformations such as singular value decomposition, principal component analysis, or independent component analysis. These methods help to identify the most important components in the multivariate signal [Birjandtalab et al., 2016; Fatichah et al., 2014]. The combination of these alternatives with classification algorithms, such as neuronal networks and support vector machine, has been successfully applied with a high sensitivity in seizure detection [Zhao et al., 2016].

Due to the reduced number of studies that perform a combination of both types of methods: spectro-temporal techniques and component analyses, it could be possible to obtain better results mixing these techniques for epilepsy seizure detection.

#### 1.1.2.3 Hypothesis

Epilepsy seizure marks can be properly detected and predicted using a one-dimension reduction of the multivariate electroencephalographic signals and efficient classification algorithms.

### 1.1.3 Method II: Nonlinear features in channel-limited EEG

#### 1.1.3.1 Problem formulation

*Is it feasible to detect seizures using the brainwave signal recorded from a reduced quantity of electrodes?*

#### 1.1.3.2 Justification

Epileptic seizures are commonly recognized due to epileptiform waves in the EEG of the patients, even they are not conclusive marks [Sierra-Marcos et al., 2015]. According to the seizure type, several zones present a higher variation in their normal working mode. Nevertheless, the whole brain presents little modifications in its behavior [Tzallas et al., 2009]. To differentiate those variations, the typical tools used to collect the signal from different positions that cover the entire head. And, consequently, more complex electrode installations are needed, and less realistic for conditions out of health facilities.

Other studies focused in specific areas to detect some types of activity, reducing the number of required sensors for recording. Using small electrode sets and nonlinear features, motor movements and eye blinking were successfully recognized. Nonlinear feature extraction has been effective for overcoming the nonlinearity of biosignal and improving the performance of classification algorithms [Tu and Sun, 2012; Spyrou et al., 2016; Natarajan et al., 2004]. Thus, with a reduced quantity of electrodes, it could be possible to recover enough information, using proper features, to detect variations in the patient condition regarding his normal state, and, consequently, detect seizures.

#### 1.1.3.3 Hypothesis

Seizures can be satisfactorily detected and predicted using the signal recorded from only two electrodes with nonlinear feature extraction.

## 1.2 Objectives

### 1.2.1 General objective

To accurately identify and classify epilepsy seizure epochs using electroencephalographic data coming from the complete set or a subset of electrodes.

### 1.2.2 Specific objectives

#### 1.2.2.1 Case study I

- To select a method to obtain the spectro-temporal information of biosignals.
- To establish a technique to perform a dimensionality reduction of multivariate electroencephalographic data.
- To select a framework to extract and to select the most important features of the one-dimensional data.
- To define a machine learning classification algorithm to be applied.

#### 1.2.2.2 Case study II

- To select a representative subset of electrodes for recognition of epilepsy seizures.
- To determine a method for preprocessing the signal while temporal and frequency variations along time are kept.
- To specify a set of the most representative linear and nonlinear features of the preprocessed signal.
- To select a machine learning classification algorithm to be employed with the brainwave features.

## 1.3 Outline

This document was composed as a collection of papers being in accordance to the regulations and recommendations of Comissão do Programa de Pós-Graduação em Ciência da Computação from Universidade Federal de Viçosa. Here, we include two papers, related to our research work, that were included in the text as separate chapters. To introduce the meaningful concepts about the field of study, we added some essential concepts in the initial chapters. The structure of this document is as follows.

Chapter 1 introduces the research context and describes the research problem, and our objectives.

Chapter 2 to 5 describe the most significant concepts about four different fields of study related to our research. Electroencephalograms as electrical biosignals, epilepsy as a neurological disease, signal processing focusing on frequency-filtering, and a machine learning overview. Each chapter explains the most relevant definitions that are useful for a better understanding of the papers.

Chapter 6 includes the first published paper. It exposes our method for epilepsy seizure detection using a conjunction of several techniques: a dimension reduction method, an envelope detector, a regrouping process, and a random forest classifier. The required input data for this algorithm were 23 brainwave signal channels. And, the overall performance of our method was better than the state-of-art alternatives.

Chapter 7 contains the second paper. It describes our second method for epilepsy classification based on the use of filter banks, nonlinear features extraction, and a decision tree. The input for this technique is the differential signal from only two electrodes. This is the main difference regarding the alternative proposed method. As it is explained in the paper content, the design conditions allowed future implementations in small devices.

Chapter 8 introduces the conclusions of the overall research and suggestions for future work.

The complete reference of both papers are as follows:

- PINTO, M. A.; CERQUEIRA, F. R. . **Patient-specific epilepsy seizure detection using random forest classification over one-dimension transformed EEG data**. *In: 2016 International Conference on Intelligent Systems Design and Applications (ISDA)*, 2016, Porto, Portugal. To be published in the Advances in Intelligent Systems and Computing proceedings Vol. 557.

- PINTO, M. A.; CERQUEIRA, F. R. . **Data mining with filter-banks for seizure detection using electroencephalograms**. Article submitted to the Electronic Letters journal.

# Chapter 2

# Electroencephalography

## 2.1 Brain electrical activity

Neurophysiology was established in the 19th century to study the relationship between the electrical waves recorded from certain muscle nerves. Later on, it was found that the brain also emits some electrical signals, and that it could be electrically stimulated [Sanei, 2013].

Neurons are nervous system cells that have the function of transmitting information through an excitable electrical membrane. [da Silva, 2009]. There are approximately between 86 and 100 billion neurons [Azevedo et al., 2009; Nidal and Malik, 2014]. For signal transmission between two neurons, a chemical process must take place in their membrane: An exchange of sodium ($Na^+$) and potassium ($K^+$) ions. This induces a change in their electrical properties. At rest, the voltage is approximately -70mV. However, during the excited oscillations, the potential ranges from -70mV to +10mV [Nidal and Malik, 2014]. These modifications have also influence on the surrounding intercellular space generating other effects such as compensatory extracellular currents [Barkley, 2004].

Over time, three techniques to measure the electrical properties of the brain were proposed:

- The first developed method, which originated the study of neurology, was the *electroencephalography* (EEG), proposed by H. Berger [Berger, 1929]. It was defined as the "recording of the electrical activity along the scalp" [Nidal and Malik, 2014].

  It should be noted that EEG only records signals from neuron groups that fulfill some conditions: They must have a synchronous activity with similar spatial orientation [Ahmad et al., 2015], and they have to be organized in macrocolumns perpendicular to the cortex surface [Akalin Acar and Makeig, 2013].

Even with its restrictions, several neurological diseases can be identified using EEGs [Al-Fahoum and Al-Fraihat, 2014].

- Other popular technique is the *magnetoencephalography* (MEG). This method differs from EEG in the kind of signal that is collected, because MEG does not directly record the electrical signals over the head, but instead, measures the magnetic field generated by the neuron paths. The resultant magnetic fields are weak, but could be recorded without interference in their route over the brain. MEG is widespread used to resolve the source localization problem.

- The third recording method is the *electrocorticography* (ECoG). It is an invasive procedure that places the electrodes on the naked surface of the brain. The absence of the skull allows to obtain signals with the lowest noise and makes it possible to recognize the source of the signals. It is useful to identify regions as potential sources of epilepsy [Leuthardt et al., 2006]. However, MEG is considered a surgery procedure and must meet proper clinic and medical requirements. This disadvantage avoids to use ECoG for non clinical applications.

Currently, the study of epilepsy seizures is separated in two groups: Behavior, and detection. In the first group, magnetoencephalograms (MEG) with LFPs and ECoGs are utilized, because they provided exact measurements with a third-dimensional view. In this way, the seizure origin and its flow patterns could be tracked across several areas of the brain. However, these methods require medical attendance with strict safety conditions. On the other side, the second group (seizure diagnosis and detection) relies on the use of EEG and video monitoring to confirm the analysis. Due to the fact that these activities must be executed in different, and non-clinical, environments, the non-invasiveness characteristic of the made of EEG a reliable measurement technique.

In this study, we focused on EEGs due to three main characteristics they present regarding another alternatives: First, EEG recordings are safety medical processes due to they are noninvasive and do not require special surgeries or specific precautions [Alotaiby et al., 2014]. Second, they have a high time resolution, because the current equipment technology allows milliseconds of sampling periods (sampling frequencies greater than 1KHz.) Thus, high-speed temporal dynamics could be

**Figure 2.1.** EEG recording from a patient with epilepsy [Goldberger et al., 2000b]

studied. And, last but not least, EEG devices are relatively inexpensive [Freeman and Quiroga, 2013].

## 2.2 Electroencephalography

EEG is the depiction of the electrical current variations along intervals of time. It was the first type of representation of brain signals (Figure 2.1). Sometimes, it is also called as "raw data" because the amplitude signal is drawn in the ordinate axis with the time as the abscisa.

Additional advantages of electroencephalography are:

- It is independent from cultural and ethnic factor, being properly a feature of the human genetic evolution [Sakkalis, 2015];
- It is "one of the most heritable characteristics in humans" [Sakkalis, 2015];

- Its oscillations are phylogenetically maintained, and they have a meaning related to cognitive functions [Berger, 1929; Sakkalis, 2015]; clusters of different sizes;
- It could be used for showing areas affected by head injuries, strokes, or areas influenced by tumours [Sanei and Chambers, 2007];
- It shows the zones from where seizures are originated, and consequently, the effects of drugs against epilepsy [Sanei and Chambers, 2007].

However, EEGs have several drawbacks: they are nonlinear and nonstationary, Thus, typical signal process techniques became ineffective, because the noise is almost as strong as the EEG signal (low signal-to-noise ratio), requiring several preprocessing techniques prior any kind of analysis. And they usually involve a high complexity analysis process [Freeman and Quiroga, 2013].

### 2.2.1 Analysis techniques

The current importance of EEG resides in its relationship with neurological diseases, psychiatric disorders, neural response to medication, and consciousness states. Thus, a proper analysis of the recordings helps to diagnose several clinical conditions [Koenig et al., 2002].

The habitual analysis of EEGs is based on the idea that the brainwave recordings must be analyzed as a summation of stationary oscillatory processes [Koenig et al., 2002]. Even though, it is known that a brain signal is not stationary. However, that "stationary" perspective helps to diagnose probable abnormalities in the health state of a patient, because the typical pattern waves and their meanings are known [Koenig et al., 2002].

A common electroencephalogram has some limitations for signal analysis: Non-repeatability, and non-periodicity. Recordings can never be exactly recreated, because the conditions will be different each time. Consequently, there is no a base pattern of a normal wave. An exhibition of some periodicity could even be a symptom of an abnormal function [Kiasaleh, 2015]. These typical wave fluctuations are due to age changes, gender differences, skull thickness, mind state, and other factors [Durka, 2007].

In this context, the neuro-specialist acquires the responsibility for determining whether a signal is a normal variation of a healthy state, or the result of a mental function impairment or the consequence of damage in some brain area [Vogel, 1970]:

> "Every experienced electroencephalographer has his or her personal approach to EEG interpretation. [...] There is an element of science and element of art in a good EEG interpretation; it is the latter that defies standardization [Durka, 2007]."

## 2.3 EEG frequency response

An electroencephalogram is usually described using its frequency components. A normal brainwave ranges from 0.1Hz to 100Hz, that means it has component waves with periods from 1ms to 1000ms. Nevertheless, the usual analyzed frequency band, which contains the relevant information, ranges from 0.3Hz to 30Hz [Ahmad et al., 2015; Miller, 2007].

Conventionally, the whole frequency band of EEG is separated in some intervals or bands, also called rhythms: Delta (0.5Hz-4Hz), theta (4Hz-8Hz), alpha (8Hz-13Hz), and gamma (35Hz-100Hz) waves [Miller, 2007; Kamel and Malik, 2015]. But it should be noted that there is no exact and clear separation between these rhythms.

The decomposition of an EEG in several rhythms is shown in Figure 2.2. The brainwave is limited to the first two seconds and belongs to the epilepsy-diagnosed patient "chb01" whose recordings are available in the dataset CHB-MIT that is described in [Shoeb and Guttag, 2010].

Brainwaves carry information through frequency variations that reflect wave changes [Miller, 2007]. There is an inherent modulation process controlled by the cognitive system that regulates these oscillations using neurotransmitters as messenger signals [da Silva, 2009].

### 2.3.1 Delta waves

The delta waves appear in the frequency band of 0.5Hz to 4Hz. They have the highest amplitude, but lower frequency in the spectrum. Usually, they can be more easily detected in the frontal side in adults and on the occipital region on the case of children.

Functionally, these waves are related with deep sleep, subcortical lesions, or metabolic encephalopathy hydrocephalus [Kamel and Malik, 2015].

It should be remembered that a constant signal has a null frequency, for this reason, it is often wrongly supposed that neuronal groups that work with a very low

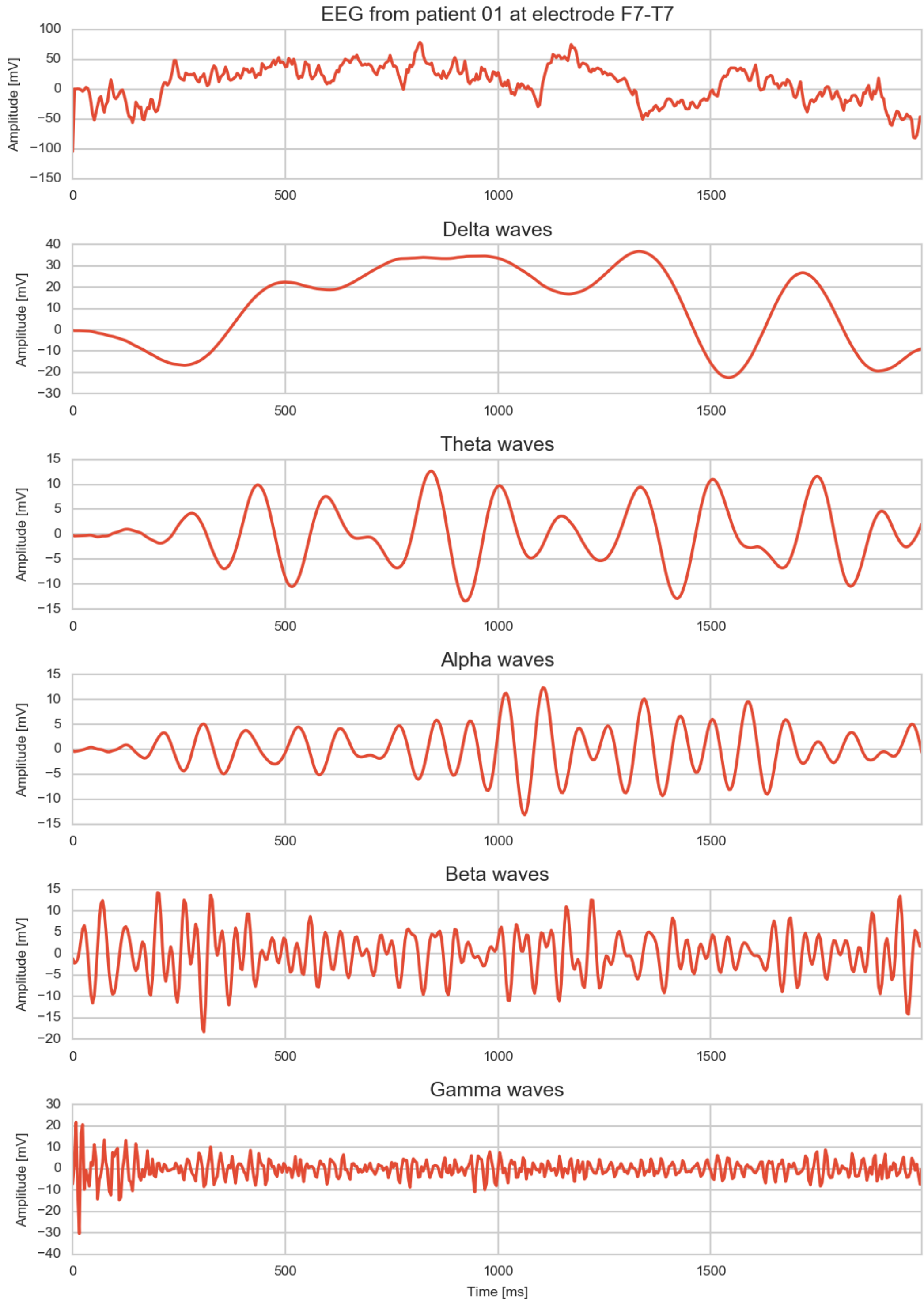



**Figure 2.2.** Brain rhythms. Patient chb01 of the CHB-MIT database [Shoeb and Guttag, 2010].

frequency have an electrical inactivity. However, some kind of cortical neurons in REM sleep or wakefulness could have an observable signal in the delta bandwidth.

### 2.3.2 Theta waves

This rhythm has a frequency interval from 4Hz to 8Hz, and it is commonly linked to young children. In adults, it could be an indicator of drowsiness or meditation, but it is not directly linked to a specific activity. An excessive quantity of these waves could be a symptom of an abnormal brain working [Kamel and Malik, 2015].

### 2.3.3 Alpha waves

These waves range from 8Hz to 13Hz. They were the first studied type of brainwaves. Initially, they were considered to be idling types of waves. However, subsequent studies showed that they are linked with several cognitive tasks such as wakefulness, alertness, creativity, and other forms of general information processing. These waves have their maximum amplitude in the occipital area [Kamel and Malik, 2015; Miller, 2007].

This kind of waves has a linkage to visual tasks. This rhythm is useful to detect eye blinking. At the same time, this type of signals is too sensitive to external factor such that an excessive illumination could reduce the amplitude of the measured signal [da Silva, 2009].

### 2.3.4 Mu waves

The mu waves are overlapped with the alpha rhythms (8Hz-13Hz), but their meaning is totally different. Mu waves are the result of synchronous action potential firing of large groups of neurons during a motor movement [Kamel and Malik, 2015].

### 2.3.5 Beta waves

This classification of signals is present in the range of 14Hz to 26Hz. Its meaning is commonly connected to active attention, active thinking, problem solving, response to visual stimulus, and sensory motor activities. These waves use to have amplitudes lower than 30uV [Kamel and Malik, 2015; Miller, 2007].

Commonly, beta band waves cannot be adequately registered with typical EEG, because they occur in specific and sparse points of the brain, which makes their resultant electrical flow unable to reach the surface [Miller, 2007].

### 2.3.6 Gamma waves

These waves lie in the band of 30Hz to 100Hz. Their functions, mechanisms, and meanings remain a matter of controversy and debate. However, they rearely appear naturally in the human brain [Kamel and Malik, 2015].

# Chapter 3

# Epilepsy and seizures

According to the World Health Organization (WHO), in 2005 more than one hundred million people worldwide suffered a mental, neurological, or a behavioral disease (Figure 3.1). Excluding the neuropathies related to nutritional deficiencies, sequelae of injuries, infections, or other non-neurological diseases, epilepsy is the most prevalent disorder, and it officially affects more than 50 million people, although it is estimated that the actual number of patients could be close to 200 million people [Tulchinsky et al., 2014; WHO, 2006].

Epilepsy is a common chronic neurological disease defined, by WHO, as a disorder of the brain characterized by a tendency to generate seizures. According to the same organization, a seizure is a transient occurrence of symptoms attributable to abnormal neuronal activity [WHO, 2006]. Thus, the diagnosis of an epileptic

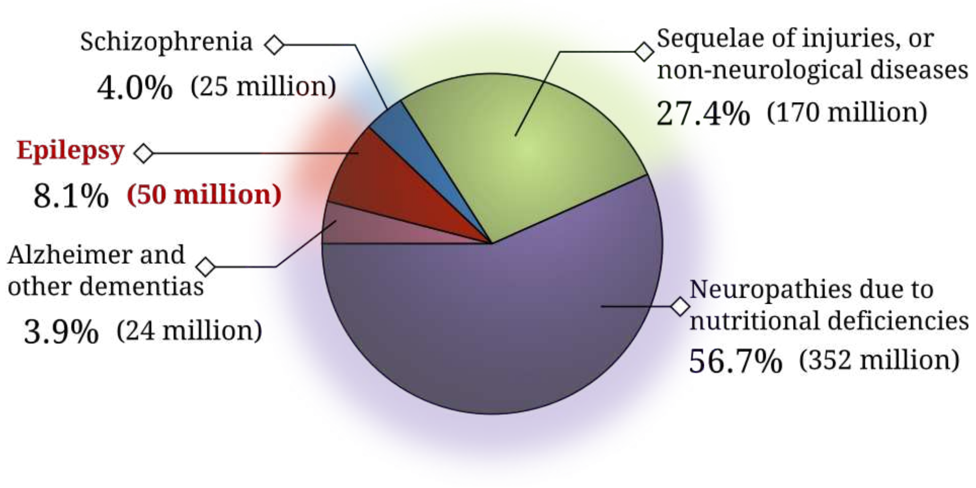



**Figure 3.1.** Global prevalence of mental illness [WHO, 2006]

condition relies on the confirmation of a predisposition to suffer seizures. However, this type of diagnosis is an unpractical situation due to subjective interpretation of a *persistent predisposition*. Consequently, a common operational definition of epilepsy is the occurrence of two or more non-provoked seizures in the interval of 24 hours [WHO, 2006].

Following these definitions, the association of seizures to epilepsy is often false, as several systemic, toxic or metabolic brain injuries lead to seizures as their symptoms in acute phases. These events are called provoked or acute symptomatic seizures. They are not repeated when the acute phase has elapsed or the illness is treated. Therefore, they cannot be considered as an indicator of underlying epilepsy [WHO, 2006]. At the same time, some types of paroxysmal activities, i.e., abnormal synchronous discharge, are incorrectly linked to seizures, and, consequently, wrongly considered as epilepsy signs [Tzallas et al., 2009].

## 3.1 Seizures and EEG

Epilepsy is a complex disease because it affects several dimensions of the patient's life: cognitive, psychological, and social circumstances. Due to the implications of suffering recurrent seizures, patients require special and constant supervision in their activities. With proper medication, almost 70% of the affected subjects have a normal life without severe restrictions. Leaving the illness untreated increases the probability of extreme complications. The number of deaths indirectly caused by an epileptic episode is 1.3 to 9.3 times greater than the number of expected deaths in similar age ranges [Tulchinsky et al., 2014; WHO, 2006].

There is a general perception that EEGs show clear and visible abnormalities during periods of seizure (cortical activation patterns known as "epileptiform waves"). In some individuals, that assumption is true because there are strong alterations in the "behavior" of their electrical patterns in one or more channels. An example of this is the epilepsy-diagnosed patient "chb01" of the dataset CHB-MIT [Shoeb and Guttag, 2010]. His recordings are shown in Figure 3.2 where there are a clear distinction between the seizure and normal states. However, that perspective of an easy distinction represents a misunderstanding of the "normality" state of a brainwave. Some patients, like the subject 'chb06" of the CHB-MIT dataset, exhibit a visually erratic pattern under in normal conditions (Figure 3.3), increasing the complexity for neurologic specialists to detect intervals of real seizures. Due to this type of cases, seizure diagnoses are aided and confirmed with long-term video monitoring [Spyrou et al., 2016].

Improving the efficiency of unprovoked epilepsy seizure detection helps to start an early treatment and, therefore, improve the quality of life and subsequent disease risks in the patient [Ahmad et al., 2015; Spyrou et al., 2016].

## 3.2 Detection methods

It is emphasized that a correct diagnosis highly depends on the subjective analysis of specialists who examine the recordings. This dependency on the clinician personnel implies a large amount of processing time due to the large size of the collected data [Tsiouris et al., 2015]. To overcome this limitation and help with the diagnosis process, several studies developed auxiliary systems that range from analog circuitry [Bhavaraju et al., 2006] to artificial intelligence-based systems [Czarnecki and Gustafsson, 2015].

We can classify the alternatives according to several characteristics:

- *Type of analyzed data*: Time series, spectral analysis, or spectro-temporal signals. Recent researches have been focused on the latter type [Zhao et al., 2016; Alotaiby et al., 2014; Das et al., 2015; Gill et al., 2015].

- *Set of extracted features*: The common approach is to extract statistical, chaos-theory or information-theory characteristics about the multivariate data [Ahmad et al., 2015; Das et al., 2015; Gill et al., 2015]. On the contrary, other studies advanced in data transformation to reduce the dataset in univariate data rather than extract features [Zhao et al., 2016].

- *The level of artificial intelligence*: It would be expected that all current methods use algorithms or procedures based on some area underlying artificial intelligence (AI). However, there are satisfactorily applied techniques that utilize only static thresholds to trigger seizure alarms [Alotaiby et al., 2014], or methods that rely on signal processing in pure analog circuitry instead of software-based processes [Bhavaraju et al., 2006].
  Effectively AI techniques usually are construct with classification algorithms such as support vector machine (SVM), artificial neural networks (ANN), k-nearest neighbors (KNN), or random forest (RF) [Orosco et al., 2013; Alotaiby et al., 2014].

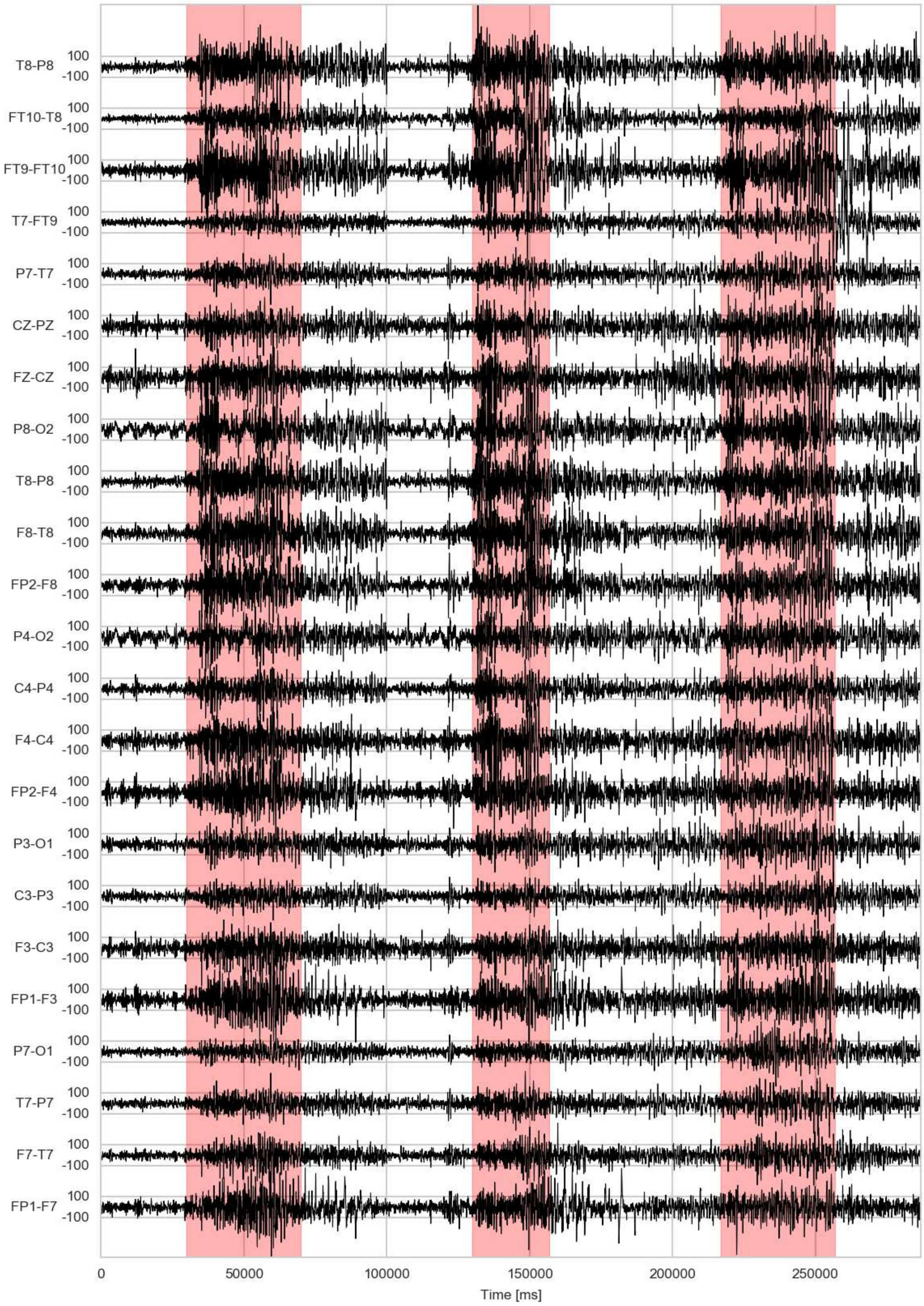



**Figure 3.2.** 23-channel EEG of patient *chb_ 01* of the CHB-MIT database with highlighted seizure periods [Goldberger et al., 2000a].

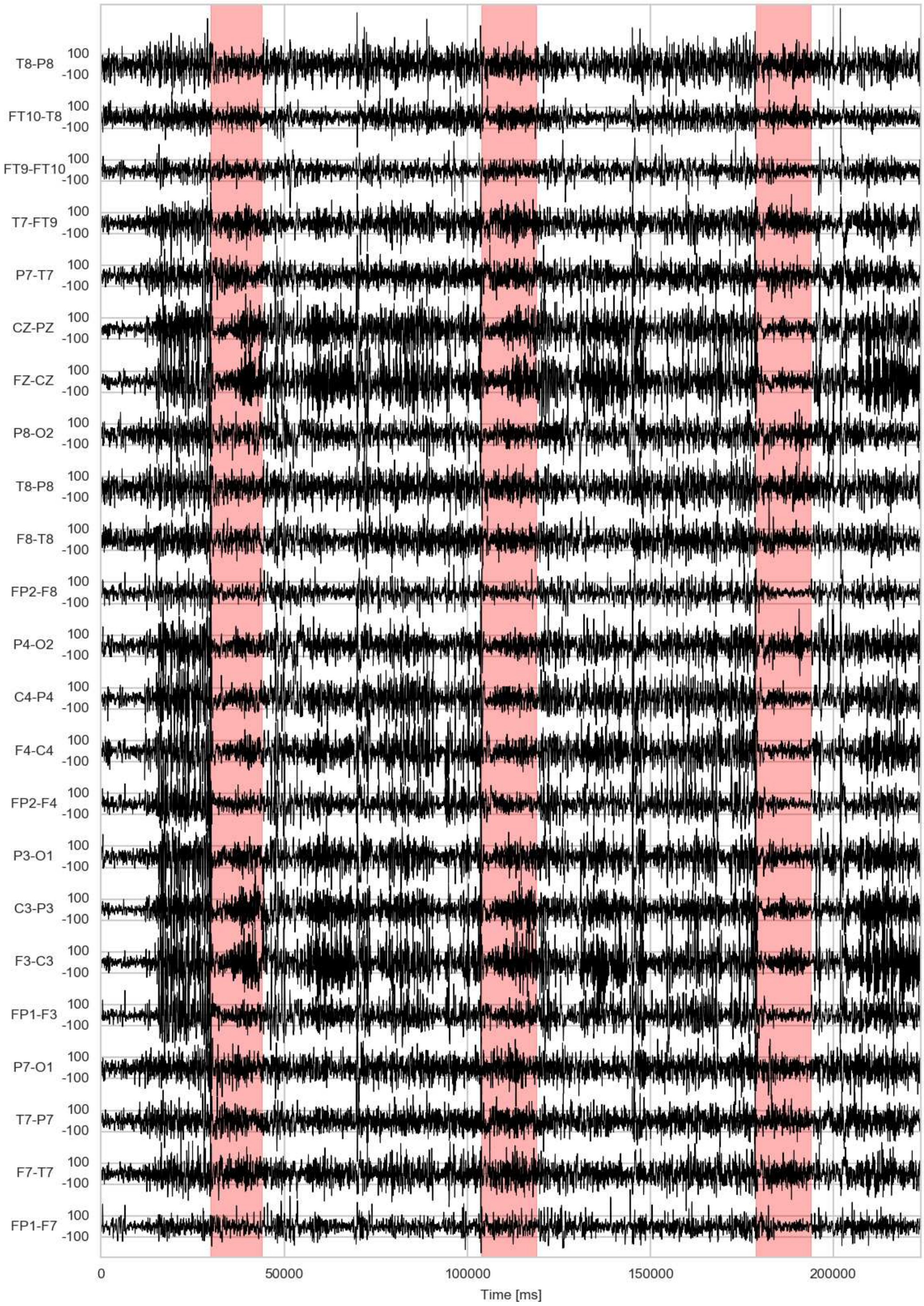



**Figure 3.3.** 23-channel EEG of patient *chb_ 06* of the CHB-MIT database with highlighted seizure periods [Goldberger et al., 2000a].

# Chapter 4

# Signal processing

This chapter introduces a brief reference for some concepts related to signal processing. As it was pointed before, we have developed two methods for detection of epilepsy seizure intervals. Both require to analyze the EEG data as signals with the appropriate methods. Moreover, and most especially, the second proposal, described in Chapter 7, highly relies on the filtering methods that are described here.

## 4.1 Signals

Signals are mathematical representations of physical phenomena. They are characterized by the information they store through variations of their parameters: amplitude, frequency, phase, energy, and so on. These changes are determined by one or more independent variables, although the common variable in nature is time [Oppenheim et al., 1998].

There are two main types of signals: discrete and continuous [Bömers, 2000]. A signal is continuous in the independent variable when time domain is infinite and completely defined in $\mathbb{R}$. Furthermore, a discrete signal is defined for only a subset $\mathbb{R}' \in \mathbb{R}$. Oppenheim uses the common notation $x(t)$ when time is continuous, and $x[t]$ when time is discrete in $\mathbb{Z}$ [Oppenheim et al., 1998]. This nomenclature is used in the equations along the text.

Pure biosignals are, in fact, continuous signals. However, for storage in digital systems, they are converted into discrete or digital signals (quantization). For instance, the original process or signal $x(t)$ is sampled at each certain time known as sampling period $T_s$, such that the new process is $x'[n] = x(n \cdot T_s)$, where $n$ is the sampling index [Oppenheim et al., 1998; Liang et al., 2013].

## 4.2 Frequency filters

A filter is a system that selectively changes several parameters of the input signal such as frequency or phase [Mandal and Asif, 2007; Bömers, 2000]. The goal of these actions is to improve the quality of the signal, removing noise or reducing the amplitude in frequency bands that does not contribute to the overall system. According to the applied technology, filters can be analog or digital. For example, biosignals are commonly frequency-filtered before a processing analysis in order to remove noise and outlier values preserving the signal of interest [Widmann et al., 2015].

Digital filter is an algorithm that operates only in a digital domain. It can be implemented as software in microcontrollers or field programmable gate arrays (FPGAs). There are two classes of digital filters: finite (FIR), and infinite impulse response (IIR) filters [Widmann et al., 2015]:

- *Finite impulse response* (FIR) filters offer a linear phase response and do not enter in resonance (permanent oscillation state). When their input is an impulse, the output is a finite quantity of non-null values. Usually, they require large amounts of memory and processing time in contrast with IIR [Dhankhar and Khaleri, 2014; Rehman et al., 2013; Aggarwal et al., 2006].

- *Infinite impulse response* (IIR) filters use a feedback working mode that optimizes memory and processing requirements of the filter. Although they offer a similar response compared to FIR with lower orders, they can have more ripples, a non-linear phase response, and resonance under several conditions [Dhankhar and Khaleri, 2014; Rehman et al., 2013; Aggarwal et al., 2006].

Filtering systems are commonly described according to several parameters in time or frequency domain. In this document, we will focus on the frequency characteristics [Laghari et al., 2014]. The frequency band that is desired to be attenuated is known as *stopband*, while the remaining and useful signal is denominated as *passband*. According to the locations of both regions, the filter is classified as low-pass (LPF), high-pass (HPF), band-pass (BPF), or band-reject (BRF) filter. Let us consider a frequency for which the amplitude is almost half the maximum amplitude of the filter. This point is called cut-off frequency $f_c$. If the filter has only one cut-off frequency and allows only frequencies greater than $f_c$, it is an HPF. Otherwise, it is LPF. However, some systems could have two $f_c$ values, if the filter allows band frequencies in the region delimited by both marks, or allows bands out of

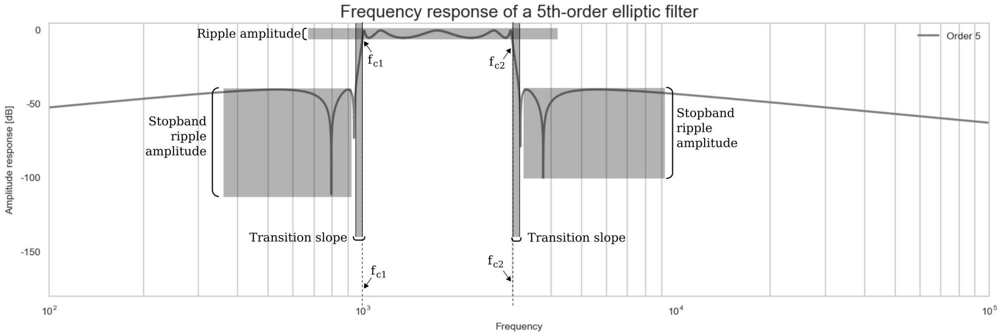



**Figure 4.1.** Typical filter design constraints

that region. These two possibilities led to filters termed BPF or BRF, respectively [Bömers, 2000].

A filter cannot automatically reduce to zero all frequencies in the stopband, because there is a natural transition from the passband, inherent to the system itself. Also, some filtering systems present some small oscillations, or ripples, in the passband or the stopband [Aggarwal et al., 2006; Bömers, 2000]. Typically, the amplitude of these ripples and the transition slope are design constraints along with the preferred amplitude of the pass and stop band (Figure 4.1).

An ideal frequency-selective filter is able to pass the specified spectral interval of the signal without ripples in any band, with a slope close to infinite, and with an amplitude of the stopband equal to zero [Widmann et al., 2015].

One of the IIR design procedures consists in adapting the classic analog methods in a digital domain, such that the knowledge of the response of that kind of filters could be directly applied to digital signals. There are three typical analog filters: elliptic, Butterworth, and Chebyshev [Dhankhar and Khaleri, 2014; Rehman et al., 2013; Aggarwal et al., 2006; Dimopoulos, 2007]. Their magnitude or amplitude response in frequency (Bode plot) is shown in Figure 4.2.

### 4.2.1 Butterworth filters

They are known as the *maximally flat response* filters [Podder et al., 2014], because they guarantee a plain amplitude with almost inexistent ripples in the passband, and with a low distortion in the phase response [Podder et al., 2014]. The frequency gain in an LPF of order $n$ is given by

$$|H(\omega)| = \frac{1}{\sqrt{1 + \left(\omega/\omega_c\right)^{2n}}} \tag{4.1}$$

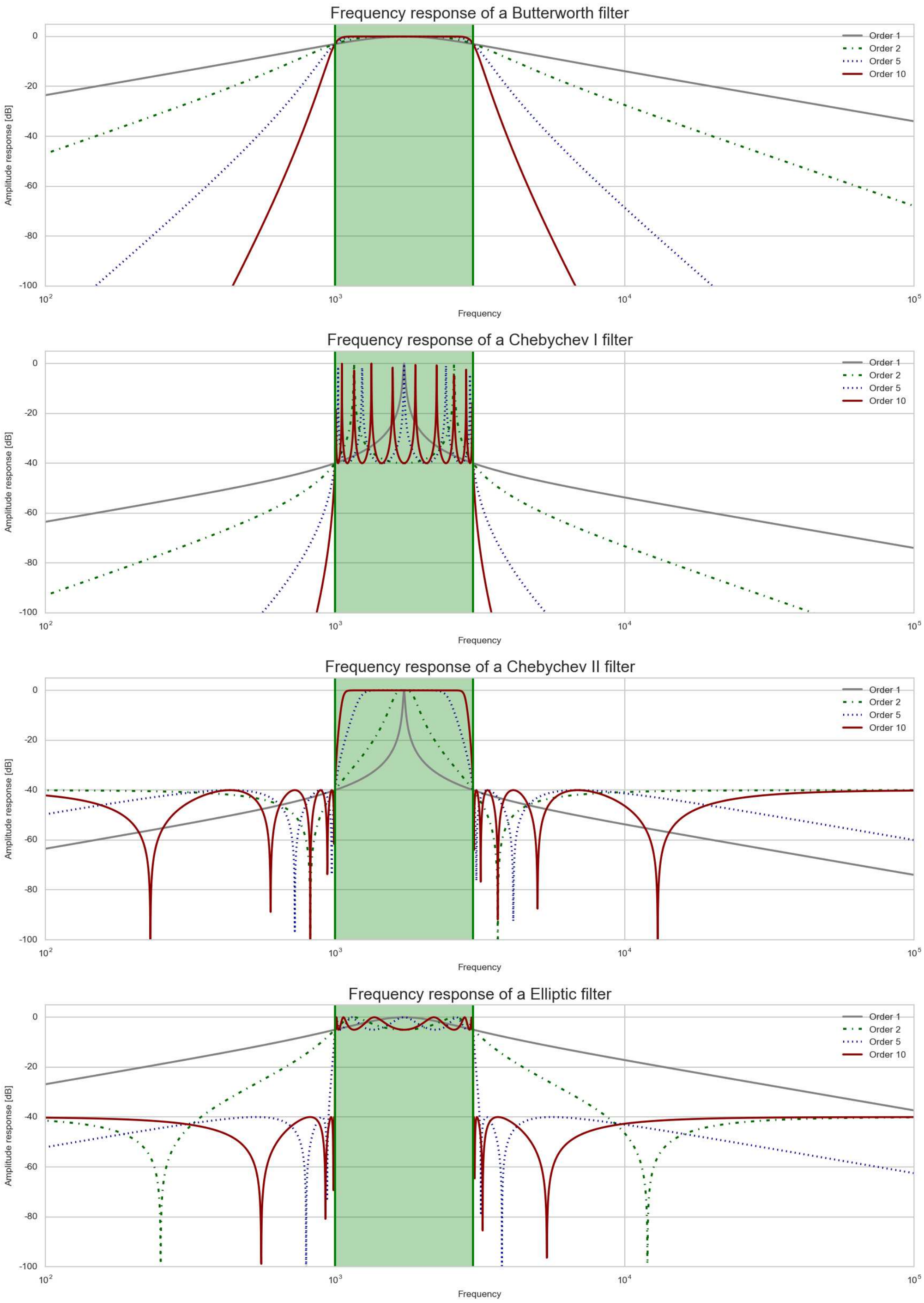



**Figure 4.2.** Magnitude response of Butterworth, Chebyshev, and Cauer filters

where $\omega_c$ is the cut-off frequency of the filter and $\omega$ the analyzed frequency. It can be noted that when the frequency tends to infinite, the output gain is zero.

### 4.2.2 Chebyshev I filters

Compared with the previous type of filter, type I Chebyshev executes a faster transition using an approximation called as *type I Chebyshev polynomial*, but with the cost of introducing ripples in the passband [Laghari et al., 2014; Podder et al., 2014; Daubechies, 1992; Rehman et al., 2013].

An Nth-order filter gain of a LPF of this type of filter is defined as

$$|H(\omega)| = \frac{1}{\sqrt{1+\varepsilon^2 T_N^2\left(\omega/\omega_p\right)}} \tag{4.2}$$

where $\omega$ is the examined frequency, $\omega_p$ the pass band corner frequency, $\varepsilon$ is the ripple control parameter that adjusts the magnitude of the ripple in the passband, and $T_N(\omega)$ is an approximation of the Chevyshev I polynomial:

$$T_N(\omega) = \begin{cases} \cos\left(N\cos^{-1}(\omega)\right) & |\omega| \leq 1 \\ \cosh\left(N\cosh^{-1}(\omega)\right) & |\omega| > 1 \end{cases} \tag{4.3}$$

### 4.2.3 Chebyshev II filters

Known as inverse Chebyshev filters, they are monotonic filters that maintain the fast transition of the previous kind of filter, but with flat amplitude in the passband, and ripples in the stopband. In applications where flatness is a requirement, this filter is preferred to type I [Laghari et al., 2014; Podder et al., 2014; Daubechies, 1992; Rehman et al., 2013].

An Nth-order filter of an LPF of this kind of filter has its gain given by

$$|H(\omega)| = \frac{1}{\sqrt{1+\frac{1}{\varepsilon^2 T_N^2\left(\omega/\omega_p\right)}}} \tag{4.4}$$

where $\omega$ is the desired analyzed frequency, $\omega_s$ the stopband corner frequency, and $T_N(\omega)$ the approximation of the Chebyshev polynomial.

### 4.2.4 Elliptic filters

Elliptic, or Cauer, filters are systems with transitions faster than the Butterworth or Chebyshev counterparts, but with ripples in the passband and stopband. They usual meet the requirements using a lower order compared to the other kind of filters [Laghari et al., 2014; Podder et al., 2014; Daubechies, 1992; Rehman et al., 2013].

The gain of a Nth-order LPF filter of this kind of filters is given by

$$|H(\omega)| = \frac{1}{\sqrt{1 + \frac{1}{\varepsilon^2 U_N^2(\omega/\omega_p)}}} \tag{4.5}$$

where $\omega$ is the frequency, $\omega_p$ the passband corner frequency, $\varepsilon$ the ripple factor, and $U_N(\omega)$ the Jacobian elliptic function or the Nth order elliptic rational function, as described in [Dimopoulos, 2007].

# Chapter 5

# Machine learning

Machine learning (ML) is a branch of artificial intelligence [Melville et al., 2009] focused on the automatic acquisition of knowledge from data and experience [Langley and Simon, 1995; Mitchell, 2006]. It is considered as a natural outgrowth of the intersection of computer science and statistics, which aims to construct algorithms on the basis of predefined structures and experience to create some inferences about data [Mitchell, 2006].

The base step in a usual machine learning process consists of learning from data: a set of objects, or instances, associated with categorized or continuous values. The following step is the evaluation of the performance of the algorithm using proper metrics. Both procedures are known as *training* and *testing* processes [Langley and Simon, 1995].

ML algorithms are often divided into two types: supervised, or unsupervised. If the algorithms depend on pre-assigned categories or labels of portions of the data to generate their models, they are known as *supervised*. Otherwise, if the algorithms do not use, or do not require *a priori* information they are named *unsupervised* [Malutan et al., 2013]. Also, the first group can have specific objectives: to classify data into specific labels or categories (classification algorithms), or to predict new values based on the past ones (regression algorithms) [Sahu et al., 2015].

ML was been successfully applied in several fields: speech recognition, computer vision, bio-surveillance, robot control, and discovering unknown patterns in empirical sciences. In those fields, it demonstrated to be an effective field for predicting and classifying data [Mitchell, 2006].

The vast majority of studies that uses electroencephalograms as data source use machine learning algorithms. Albeit nomenclature varies from one author to another, the typical process of a prediction system includes three steps [Ahmad et al., 2015; Kotsiantis et al., 2006]:

- **Data preparation and pre-processing**. The performance of the algorithms strongly depends on the quality of the data used for training [Napflin et al., 2007]. Thus, the first stage of any classification is to clean data, repair missing values, remove probable outliers, normalize data. Also, in the case of continuous time series, it is common to apply a frequency filter to reduce noise values [Alotaiby et al., 2015; Lee et al., 2014; Palaniappan, 2006].

- **Feature extraction**. A feature is a distinctive or characteristic measurement, transformation, or component obtained from the data [Al-Fahoum and Al-Fraihat, 2014]. Then, the original data is transformed into a set of features representing each data instance [Ahmad et al., 2015]. In EEG studies, the feature vector includes statistical summary measures, information-theory parameters, and other specialized features.

- **Feature reduction or selection**. Not only the quality of the data, or their features, could influence in the performance of an algorithm, but the dimension order as well: the problems caused by high-dimensional data are known as "curse of dimensionality" [Verleysen et al., 2014; Sanei and Chambers, 2007; Ahmad et al., 2015; Czarnecki and Gustafsson, 2015]. Thus, feature extraction is the process of selecting the properties with the most significant representation of the dataset. Several researches using EEG frequently employ Principal components analysis, discriminant analysis, and independent component analysis for this purpose [Al-Fahoum and Al-Fraihat, 2014; Ahmad et al., 2015].

- **Classification**. Given a set of feature vectors $X = \{X_i\}$ and a set of labels $Y = \{y_i\}$, a classification algorithm learns a predictor relation $f : X \to Y$ between the features and labels. The built model predicts the label of an instance based on its feature vector $y_i = f(X_i)$. [Mitchell, 2006]. Brainwaves studies typically rely on three types of ML classification algorithms: Artificial neural networks, support vector machine, and ensemble algorithms [Mitchell, 2006; Czarnecki and Gustafsson, 2015].

## 5.1 Principal component analysis

Principal component analysis (PCA) is a technique that linearly transforms the original variables into other uncorrelated ones. The resultant variables, known as principal components, are orthogonal among them. The PCA transformation is performed in such a way that the first principal component accounts for the highest variability in the data, the second principal component accounts for the second highest score, and so on [Patel et al., 2015]. Therefore, to all components are assigned a score denoting their variance [Czarnecki and Gustafsson, 2015]. In this way, the components with the highest scores are kept to reduce the data dimensionality.

In [Zhao et al., 2010], it was studied the influence of different mapping functions, or kernels, and PCA along with a support vector machine classifier with EEG data. It was shown that three dimensions could reach up to 100% of accuracy in several conditions concerning the classifier settings, the type of mapping function, and the kind of mental state analyzed.

## 5.2 Classification

Classification can be defined as the process of prediction a category $y_k$ based on a set of features $X_k = (x_1, x_2, \ldots, x_n)$. This prediction needs a previous learning process where the algorithm is trained with sets of features, with labels respective, to infer a data model [Ahmad et al., 2015; Czarnecki and Gustafsson, 2015].

As Wolpert et al. cited as the “no free lunch” (NFL) theorem, there is no machine learning algorithm that performs well for every problem [Wolpert, 1996]. Thus, according to the characteristics of the datasets and their features, each algorithm will have a different prediction performance: Support vector machines work very well for prediction of emotion in texts [Martinez and Martin, 2011], random forest for visual identification of chromosomes [Saranya et al., 2015; Roget et al., 2016], and k-neighbors for malware detection [Urcuqui and Navarro, 2016].

Regarding the performance evaluation, the usual evaluation metrics are based on binary classification, where the labels are positive or negatives. The trained algorithm is applied to a testing set to measure its performance. When the predicted label is identical to the actual one, the result is named as true positive (TP) or true negative (TN), if the label is positive or negative, respectively. Otherwise, a wrongly predicted negative is called false negative (FN), or false positive (FP) if it was predicted with this label (Figure 5.1). To quantitatively measure the performance, it is common to use several parameters [Orosco et al., 2016; Das et al., 2016; Tsiouris et al., 2015; Ahmad et al., 2015; Gill et al., 2015; Zhao et al., 2016; Iqbal et al., 2015; Urcuqui and Navarro, 2016]: Accuracy, precision, recall (or sensitivity or true

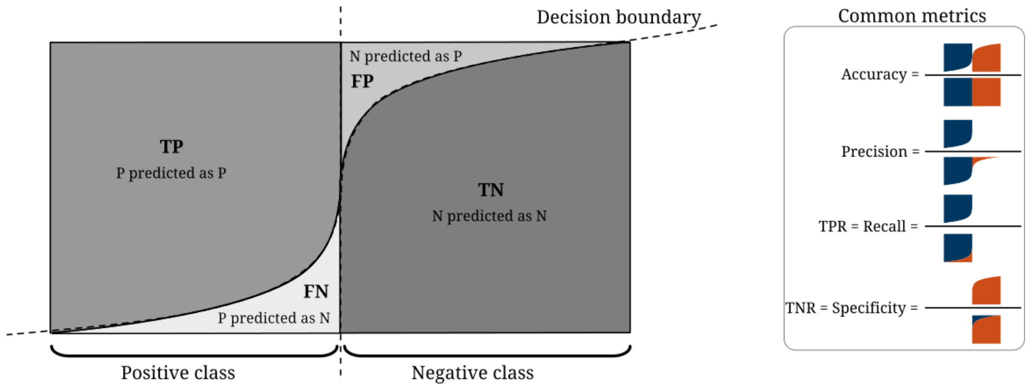



**Figure 5.1.** Performance metrics for ML algorithms

positive rate TPR), and specificity (or true negative rate TNR) defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{5.1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.2}$$

$$\text{TPR} = \frac{TP}{TP + FN} \tag{5.3}$$

$$\text{TNR} = \frac{TN}{FP + TN} \tag{5.4}$$

Below, we briefly describe the machine learning algorithms used in this work.

## 5.3 Decision trees

A decision tree is a hierarchical representation of a series of rules, where each leaf represents a classification decision, and each internal node is an evaluation rule [Melville et al., 2009]. The training process of this class of algorithms consists in creating splits of the training instances in several groups. Each split group does not require to be equally sized. It is one of the most common alternatives for different types of data due to its structure and easy interpretation [Peker et al., 2015].

In the present work, we have used a common type of decision tree: CART. CART, or "Classification and regression tree," is a non-parametric binary recursive partitioning procedure for predicting variables using categorical and/or continuous predictor variables [Breiman et al., 1984; Lawrence and Wright, 2001; Razi and Athappilly, 2005].

CART splits the current dataset into two subsets in which the target variable is more homogeneous than the original set. This process is recursively executed until it satisfies a stop criterion. In continuous variables, the condition has the form "attribute $X_i \leq C$." And, it should be noted that the same predictor variable can be used for several nodes [Speybroeck, 2012; Lawrence and Wright, 2001; Razi and Athappilly, 2005].

There are several advantages for using CART: It is a flexible method, it does not require setting parameters, it does not assume any statistical distribution of the data, and it works with any type of data. Nevertheless, CART generates a binary tree with many levels, and it does not provide any information about confidence intervals about the data [Lewis, 2000].

The key part in CART is the splitting method. CART selects a node such that the new two split subsets have less impurity than the parent subset. The data impurity is measured with, typically, four measurements: Gini, entropy, "twoing", or aggregation variance. Among them, the most common metric is Gini index [Speybroeck, 2012; Lawrence and Wright, 2001; Loh, 2011].

In the case of a binary target, the Gini measure of impurity of a node t is [Wu et al., 2008]:

$$G(t) = 1 - p(t_+)^2 - p(t_-)^2 \tag{5.5}$$

where $p(t_+)$ and $p(t_-)$ are relative frequencies of the positive class "$+$" and the negative class "-", respectively.

Considering that a node $t$ splits the data, using a condition $s$, into two nodes: $t_L$ and $t_R$, the Gini gain, or improvement, is [Wu et al., 2008]:

$$I_s(t) = G(t_L) - p(t_L)G(t_L) - p(t_R)G(t_R) \tag{5.6}$$

where $p(t_L)$ and $p(t_R)$ are the proportions of the size of each subset ($t_L$ and $t_R$) in relation to the size of $t$, and $G(t_L)$ and $G(t_R)$ are the respective impurities of $t_L$ and $t_R$.

In each partition, among the all possible candidates $S$, CART choose the best condition $s$ to maximize the improvement $I(t)$ [Wu et al., 2008]:

$$I(s,t) = \max_{s \in S} I_s(t) \tag{5.7}$$

The CART algorithm stops when or all the target values are the same ($G(t) = 0$), or the tree has reached the maximum allowed depth, or the number of the samples in the node is less than the preset minimum [Speybroeck, 2012; Lawrence and Wright, 2001; Loh, 2011].

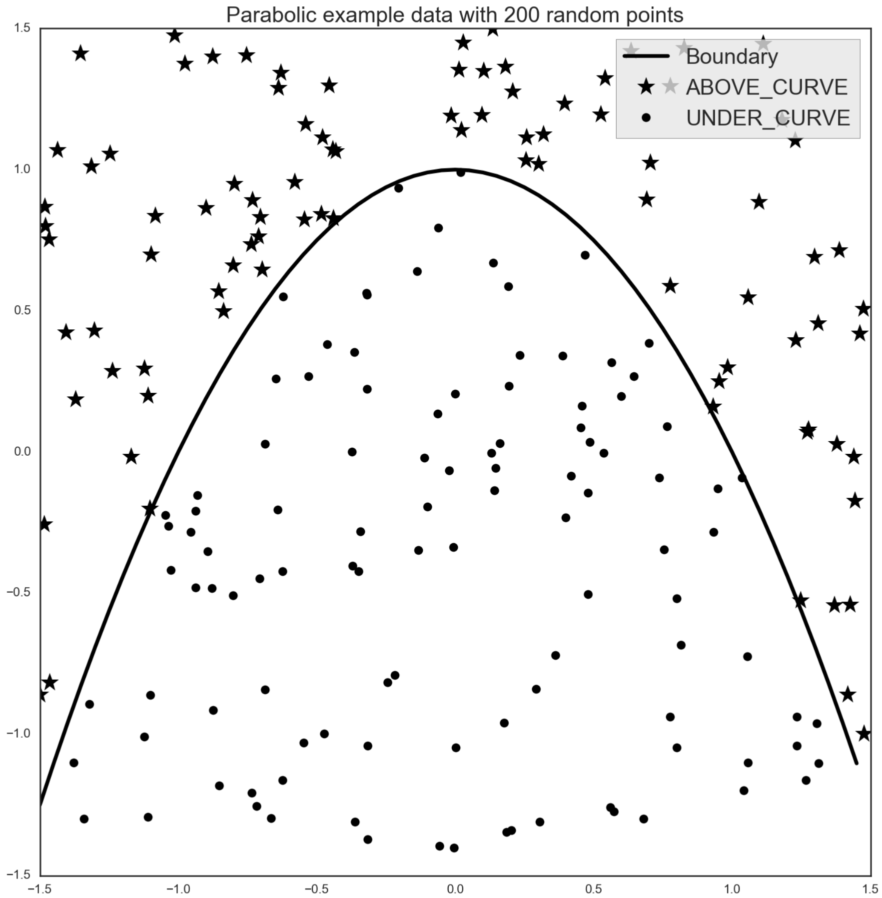



**Figure 5.2.** Non linearly separable dataset.

To appreciate the performance of this algorithm, we tested it in a non linearly separable dataset. The dataset created for this purpose has two continuous features: $x$ and $y$. For the instances, 200 points were randomly generated in the interval $(-1.5, 1.5)$. And, the class of each instance was defined according to this rule:

$$\text{class} = \begin{cases} \text{ABOVE\_CURVE} & y \geq 1 - x^2 \\ \text{UNDER\_CURVE} & y < 1 - x^2 \end{cases} \tag{5.8}$$

The final dataset is shown in Figure 5.2, where the curve equation is emphasized with a black line.

The algorithm implementation for the testing relies on the Scikit-learn 0.17.1 library of the Python language programming, due to the same library is used in this study. The testing method was splitting the entire dataset into two parts, where 60

Also, for emphasizing the importance of the pruning, we adjust the maximum depth of the CART algorithm in three values: 1, 4, and infinite. The results of this

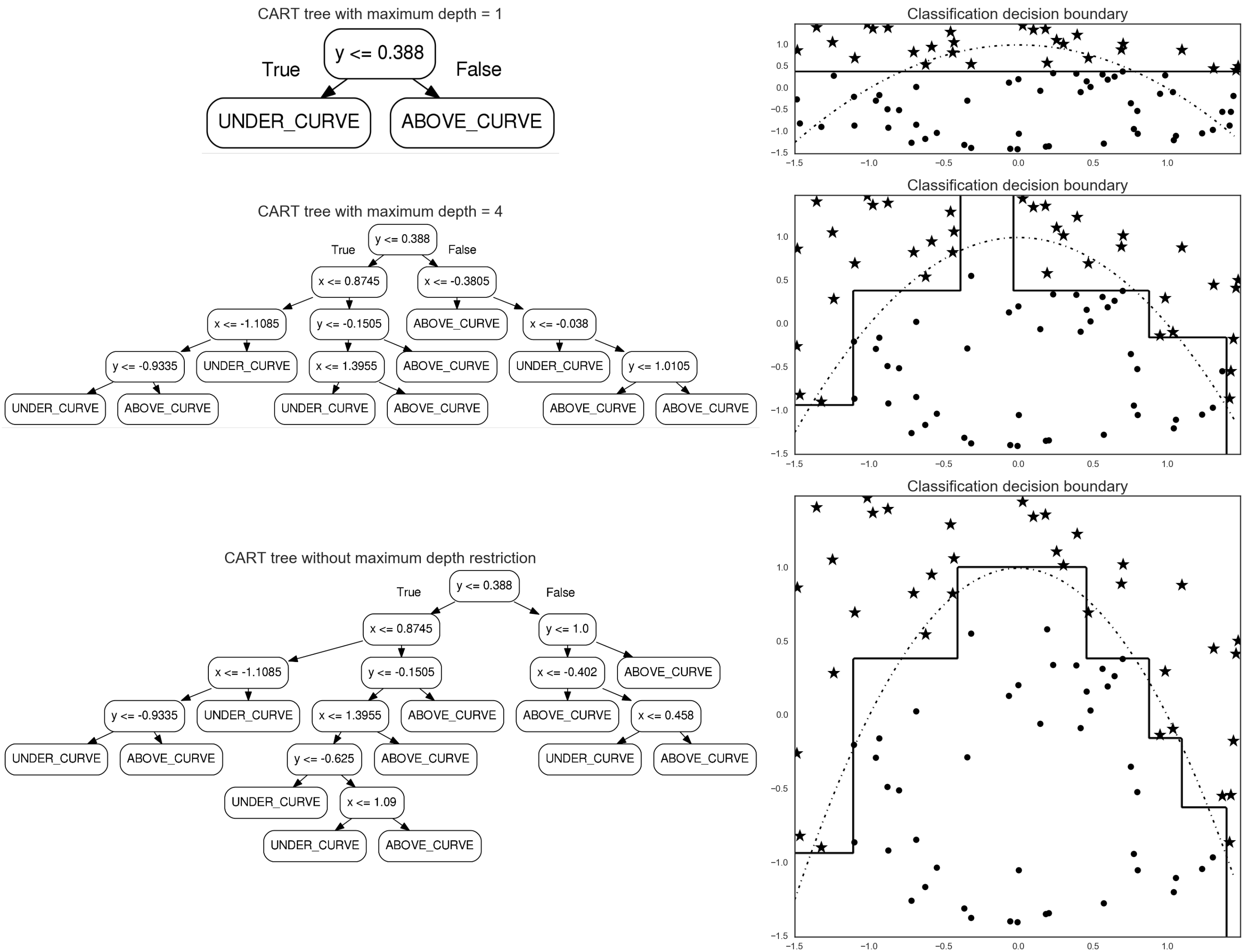



**Figure 5.3.** Example of CART algorithm over a non linearly separable dataset.

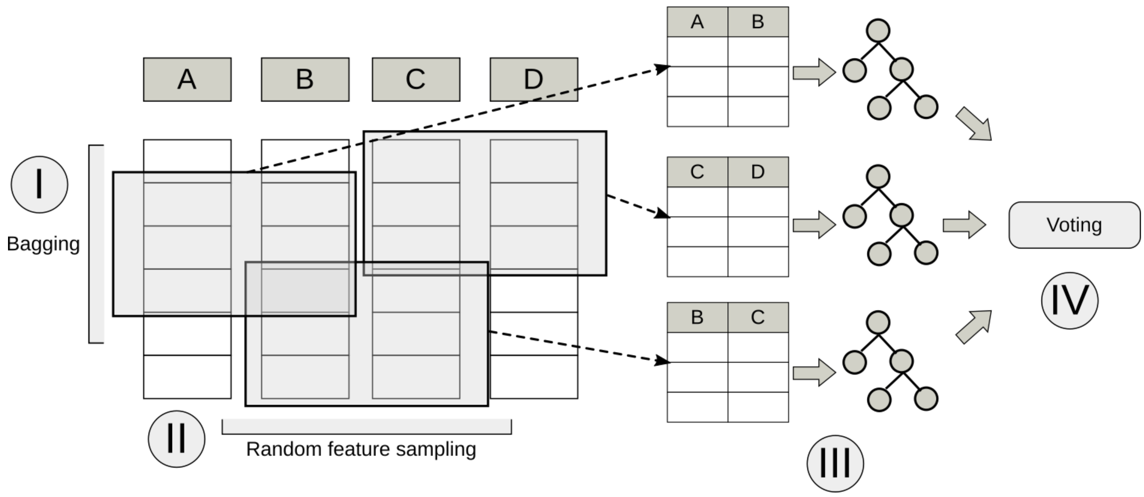



**Figure 5.4.** Procedure of a random forest algorithm.

procedure are shown in Figure 5.3. Three different trees were created, and each one with its own decision boundary. The goal of this example is to show the relationship between the depth of the tree (and its complexity), and the approximation of the classifier to the original boundary. As the tree grows, it is closer to the proposed curve, but it increases its probabilities of overfitting the data.

## 5.4 Random forest

Random forest (RF) is an ensemble learning algorithm that uses multiple decision trees [Breiman, 2001; Shiratori et al., 2015]. This method has shown to be competitive with the other classifier types in a large kind of datasets from protein-ligation classification to visual screening decision [De Bock et al., 2010; Melville et al., 2009].

The RF algorithm starts creating a new training set with random instances of the original dataset (process I in Figure 5.4.). This random selection is known as "bagging" [Chan and Paelinckx, 2008; Pal, 2005].

Later, a feature subset of size $m$, typically $m = \sqrt(n)$, is selected from all attributes of the dataset (step II in Figure 5.4) [Czarnecki and Gustafsson, 2015; Chan and Paelinckx, 2008; Pal, 2005] This subset is randomly chosen too, and this process is known as "random feature sampling" [Deng et al., 2013], or "random subspace" [Ham et al., 2005], or "attribute bagging" [De Bock et al., 2010]

With these random attributes and random sampling data, a decision tree is training looking for best separation of the data (process III in Figure 5.4) [Deng et al., 2013; Chan and Paelinckx, 2008; Pal, 2005]. Like CART, the typical metric is the Gini measure of impurity, and consequently, the Gini gain is employed to determine the best splits [Chan and Paelinckx, 2008].

The previous processes are repeated $M$ times such that the forest is comprised of $M$ trees. When a new instance is going to be evaluated, it is analyzed by all the trees, and the predicted class is decided based on the majority vote of the tree set (process IV in the Figure 5.4) [Chan and Paelinckx, 2008; Pal, 2005].

There are two parameters that are typically calibrated in an RF method: the number of trees or estimators, and the maximum depth of each tree[Svetnik et al., 2004; Pal, 2005]. To illustrate the relevance of both parameters, we considered the non-linearly separable dataset described in the previous section (Figure 5.4.) Using the same libraries, we considered random forests with 5, 50, and 500 estimators, both of them with and without a restriction of maximum depth of 3 nodes.

The random forests with a depth limit are shown in Figure 5.5. The first two columns show two examples of the generated random trees, and the last column illustrates the decision boundary reached with the entire forest. It is concluded from the graphic, that increasing the number of trees helps to smooth the shape of the decision boundaries of the RF technique, but due to the depth restriction, it was not possible to reach the expected boundary.

The three results of the same RFs without depth restrictions are plotted in Figure 5.6. As it was expected, as the number of estimators increases, the performance of the algorithm does too. And, without pruning constraints, the tree forest with 500 estimators almost reaches the shape of the equation 5.8.

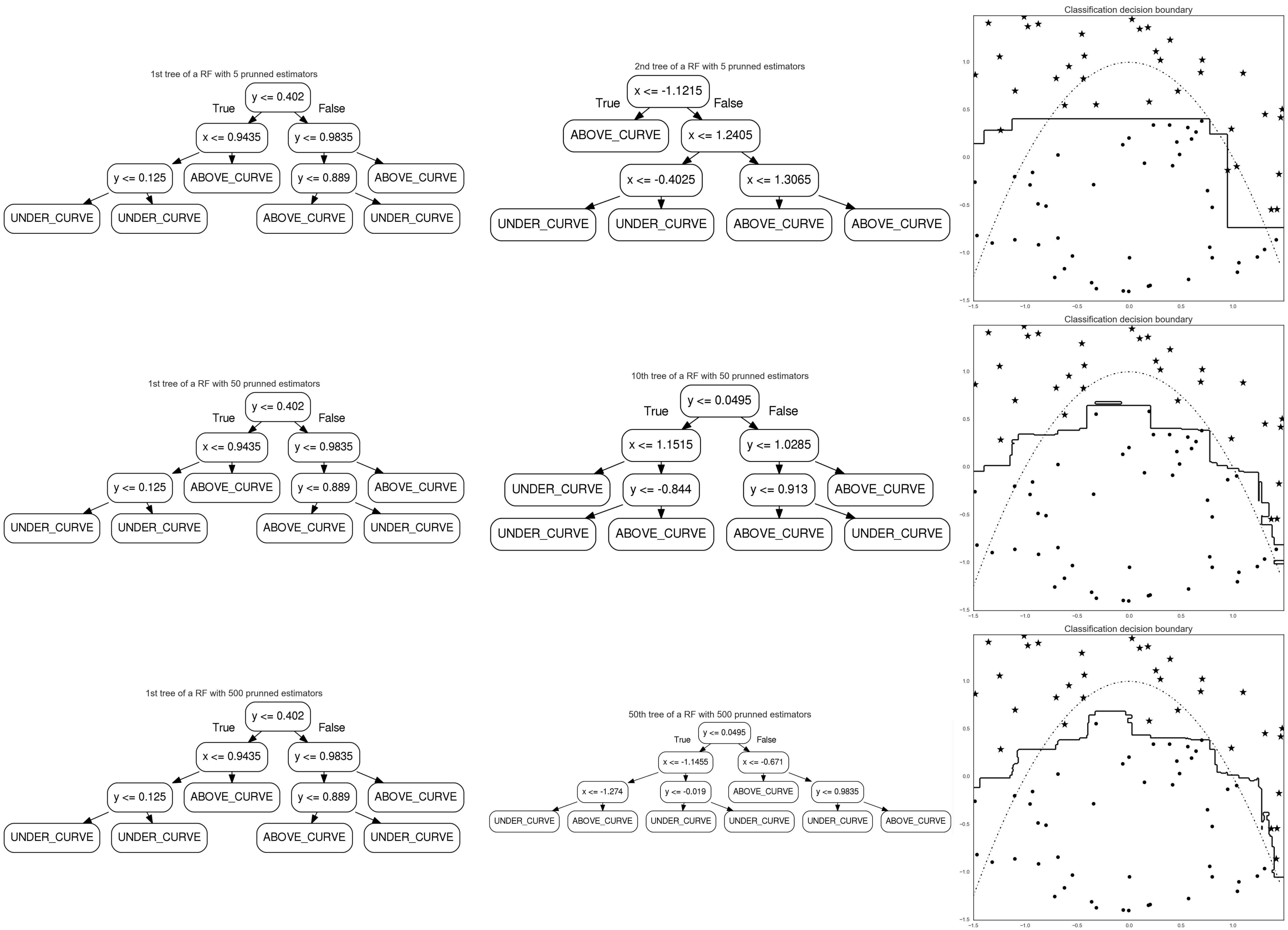



**Figure 5.5.** Example of pruned random forests over a non linearly separable dataset.

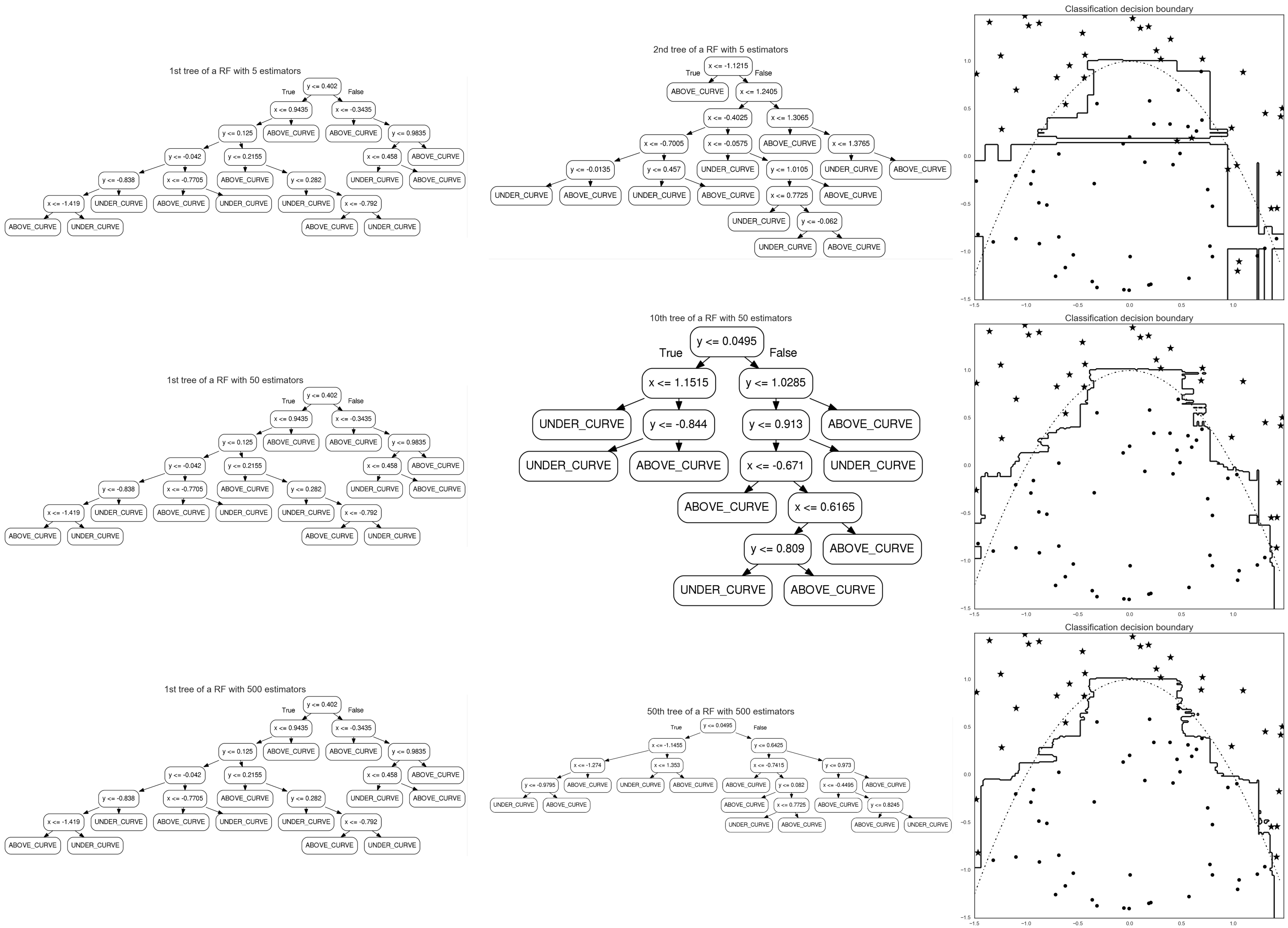



**Figure 5.6.** Example of unpruned random forests over a non linearly separable dataset.

# Chapter 6

# Patient-specific epilepsy seizure detection using random forest classification over one-dimension transformed EEG data


## Abstract

This work presents a computational method for improving seizure detection for epilepsy diagnosis. Epilepsy is the second most common neurological disease. It impacts between 40 and 50 million of patients in the world. However, epilepsy diagnosis using electroencephalographic signals implies a long and expensive process involving medical specialists. The proposed system is a patient-specific system which performs an automatic detection of seizures from brainwaves applying a random forest classifier. Features are extracted using one-dimension reduced information from a spectro-temporal transformation of biosignals that pass through an envelope detector. The performance of the present method reached 97.12% of specificity, 99.29% of sensitivity, and a $0.77h^{-1}$ false positive rate. Therefore, the method hereby proposed has great potential for diagnosis support in clinical environments.


## 6.1 Introduction

Epilepsy is the second most common neurological disease in humans after stroke [Guo et al., 2010]. It is estimated between 40 and 50 million of patients suffering from this condition worldwide, which means 1% of the total world population [Tzallas et al., 2009; Guo et al., 2010; Fatichah et al., 2014]. Although there are several medical treatments, 30% of the patient population have not a positive response to medication [Orosco et al., 2016] and require expensive and long diagnosis processes [Das et al., 2016].

Seizures are the typical indicators for epilepsy diagnosis [Tzallas et al., 2009]. A seizure is an abnormal excessive and hyper-synchronized neural activity in the brain [Tzallas et al., 2009; Das et al., 2016; Orosco et al., 2016], and it can be seen through electrical variations recorded from the whole brain mass or specific sections on the brain structure [Sierra-Marcos et al., 2015].

Among several imaging techniques of biosignals with medical relevance [Teplan, 2002], electroencephalograms (EEGs) are some of the most relevant alternatives. EEGs are recordings of electrical time-dependent variations of the brain activity [Teplan, 2002; Djemili et al., 2016]. Although the electrical amplitude measurable in each neuron is small, the result of the synchronization in time and phase of signals from large neural networks during cognitive activities allows detecting the neuron electrical signals even at scalp with EEG [Hesse et al., 2003]. It is a safe, non-invasive, and easy to assemble method [Hu et al., 2011]. Those advantages established EEG as the *de facto* standard, along with video monitoring, for epilepsy seizure diagnosis [Page et al., 2015].

Medical recommendations for seizure diagnosis often include performing long brain activity recordings of patients [Tsiouris et al., 2015]. Then, the obtained registers are analyzed by experts in the area relying on visual and subjective inspection [Orosco et al., 2016; Das et al., 2016]. This human dependency implies an expensive and time-consuming process prone to errors due to the stored data size [Tsiouris et al., 2015].

It is important to notice that not all epileptiform waves appear during seizure intervals [Sierra-Marcos et al., 2015]. Paroxysmal activity is an abnormal synchronous discharge of a large ensemble of neurons and is strongly related to seizure processes. Such an activity can be detected in EEG and is usually confused with an epilepsy marker [Tzallas et al., 2009].

Several alternatives have been developed to recognize seizures in EEG [Orosco et al., 2013; Alotaiby et al., 2014]. A first distinction among the approaches is the specific domain of the EEG data on which each method focuses: Time series, frequency data, or spectro-temporal signals [Ahmad et al., 2015; Das et al., 2015]. Another significant difference is the set of features that each method applies in the analysis. Some studies utilized statistics, chaos theory, or information theory parameters [Ahmad et al., 2015; Das et al., 2015; Gill et al., 2015], while other efforts applied data transformations such as singular value decomposition (SVD) or principal component analysis (PCA) [Zhao et al., 2016]. Furthermore, support vector machines (SVM), k-nearest neighbours (KNN), and artificial neural networks (ANN) are typically used as classifiers [Orosco et al., 2013; Alotaiby et al., 2014].

This study proposes a novel approach that presents a higher performance compared with other state of the art studies. The method herein described is a combination of several techniques of signal processing and machine learning: Short time Fourier transform, principal component analysis, maximum moving filter, and a random forest classifier.

## 6.2 Materials

In this research, the data source relies on the CHB-MIT electroencephalographic scalp database from the Physionet project[Goldberger et al., 2000b]. This database stores epileptic seizures from several pediatric patients at the Children's Hospital Boston [Shoeb et al., 2004; Shoeb and Guttag, 2010]. The data were collected for an experiment conducted for monitoring patients after withdrawal of the epileptic medication before a surgical intervention. A total of 24 datasets were built from 23 different subjects. Each dataset was labeled with a code from *chb01* to *chb24*. It should be noted that datasets *chb02* and *chb24* represent biosignals of the same patient but with a difference of 1.5 years between the recording dates. However, they are identified as different subjects in our experiments.

The complete data package is compounded of 686 files saved in the European data format (EDF) representing a total of 961.64 hours of recording [Goldberger et al., 2000b]. The seizure recording durations of the patients are in the range of 6 seconds (subject *chb16*) to 752 seconds (subject *chb11*). All signals were measured with a sampling frequency of 256 Hz with 16 bits of resolution. Nearly, all files were captured in 23 channels using the international 10-20 system of electrode positions and nomenclature.

## 6.3 Method

This study proposes to use a mapping of a spectro-temporal transform of the brain signals into a one-dimension space for being used as input to a classifier algorithm (Figure 6.1). As a result, the complete signal analysis could be explained using four

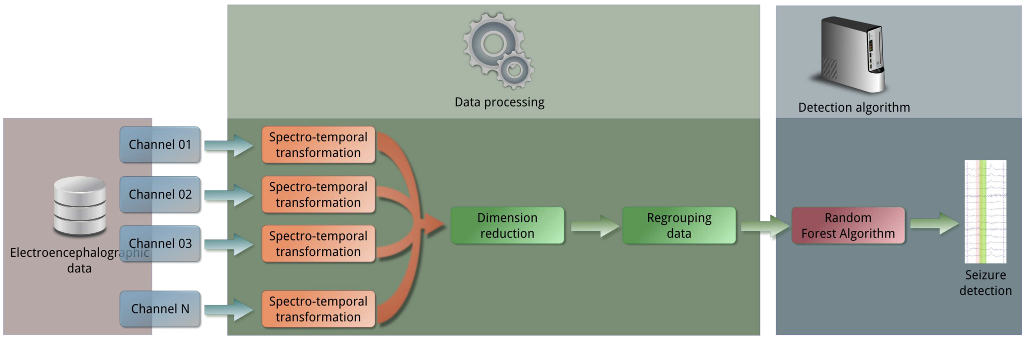



**Figure 6.1.** Workflow of the seizure detection process: Data preprocessing, dimensionality reduction, envelope detection, data regrouping, and classification.

processes: Data preprocessing, dimensionality reduction, envelope detection, data regrouping, and classification.

### 6.3.1 EEG data processing

Some characteristics are important to be explained about EEG signals. They exhibit several unique features: Nonstationary, nonlinear, and frequency-variant. The nonstationary property denotes that there is no period of time where a given signal is repeated [Natarajan et al., 2004; Yan et al., 2015]. Also, previous studies proved that variations in the information source on different conditions or activities are linked with alterations in the spectral response of the brain biosignals [Yan et al., 2015]. For this reason, the deviations that appear in the frequency domain are useful to be analyzed. The conventional alternative is a discrete Fourier transform or a fast Fourier transform [Alotaiby et al., 2014]. However, those approaches are focused in stationary data series and cannot be directly applied in EEG [Natarajan et al., 2004].

A typical approach to analyze nonstationary signals is splitting them into several time intervals, or time windows, such that spectral features can be obtained, while partial time information is retained [Tzallas et al., 2009]. This method, known as short time Fourier transform (STFT), showed effectiveness to detect seizures [Samiee et al., 2015; Yan et al., 2015] . Therefore, this mathematical tool was selected as the procedure to map EEG signals into a time-frequency domain [Wang et al., 2015]. For a given discrete function $x[n]$, STFT is calculated shifting a small sliding window $g[n]$ of size $\Delta t$ over the time series to obtain the frequency spectrum in each time interval:

$$G\left\{x\left[n\right]\right\}\left[m,f\right]=\sum_{n=-\infty}^{\infty}x\left[n\right]g\left[n-m\right]e^{-j2\pi fn} \tag{6.1}$$

where $n$ and $m$ are sampling instants, and $f$ is the signal frequency.

Due to the fact that the signal is not continuous, i.e., it is sampled at fixed periods with a frequency $f_s$, there are several limitations in the process of STFT calculation. Therefore, some constraints must be considered, because the time resolution $\Delta t$ and frequency resolution $\Delta f$ are related to the number of points $\Delta n$, as can be seen below:

$$\Delta t=\Delta n\cdot f_s \tag{6.2}$$

$$\Delta f=\frac{f_s}{2\Delta n} \tag{6.3}$$

Thus, for the datasets described previously, a time window $\Delta t=39$ microseconds was used, which allowed a range of 10 frequency intervals with a resolution $\Delta f=12.8Hz$

### 6.3.2 Data dimensionality reduction

Electroencephalograms are multivariate data and their computed values present a high dimensionality, 230 features per instance in this research, which complicates data visualization and processing [Birjandtalab et al., 2016]. This problem, known as "the curse of dimensionality", can impact the classifier performance increasing misinterpretations or overinterpretation of the information [Ahmad et al., 2015]. Common procedures to reduce these inconveniences rely on selecting the most significant features. In this work, we chose principal component analysis for feature selection, as described in [Alotaiby et al., 2014].

PCA was applied over the EEG multivariate time series maintaining only the component with maximum score. The aim of mapping to only one dimension is allowing to represent graphically the whole brain signal in one chart maintaining a relationship between the signal shape and the EEG epilepsy state.

### 6.3.3 Signal envelope detection and regrouping data

Previous experiments (not shown) have revealed several associations between the PCA reduced data and the seizure marks. This connection is apparently related

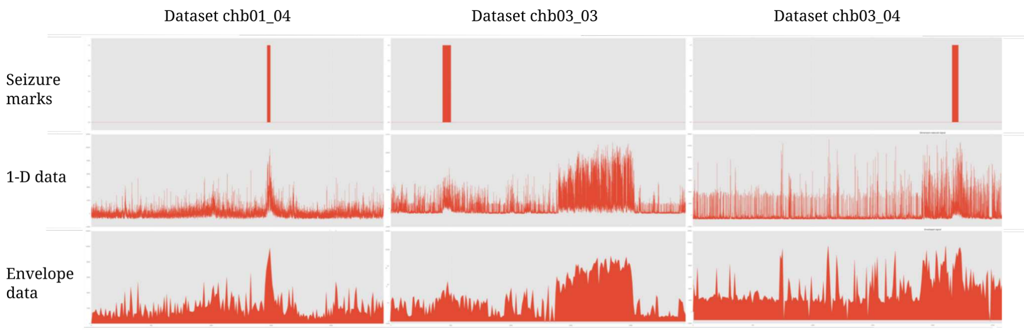



**Figure 6.2.** Graphical comparison between the PCA converted data and its envelope regarding the seizure marks in the datasets *chb01_04*, *chb03_03* and *chb03_04*.

to the PCA signal shape, as can be seen in Figure 6.2. It should be noted that this behavior corresponds to amplitude modulated signals. As a consequence, it was applied an envelope detection process over the PCA signal. There are several methods available for this goal: Hilbert transform, pass band filters, and hysteresis transform. However, after a trade-off analysis between processing time and resultant shape, the simplest off-line method was selected: The moving maxima filter.

The moving maxima filter process consists of creating a new series $h[n]$ from an original series $g[n]$, such that each point n represents the maximum value between $g[n]$ and $g[n+\Delta n]$:

$$h[n] = h'[n\Delta n] = \max_{n \leq i \leq n+\Delta n} \{g[i]\} \tag{6.4}$$

The goal of this filter, proposed by Zhang and Smith [Zhang and Smith, 2001], is finding the envelope of the signal using amechanism similar to an amplitude demodulator, which depends only on the time period $\Delta n$, defined as 10 seconds in this work.

After this process, the enveloped data were reorganized in blocks of fixed size $m$. This size may influence the final system ability to detect a seizure state and it should be tested with several values. However, due to the limitations related to the data size, only two block sizes were tested: 30 and 70 seconds.

### 6.3.4 Machine learning: Balancing and classification

The last step in our method is to apply a machine learning algorithm. However, EEG datasets are typically imbalanced [Tsiouris et al., 2015], i.e., the points marked as seizure comprise a very small set in comparison to the interictal intervals. As a result, firstly, the SMOTE algorithm was applied over the data for treating the imbalance problem. In [Fergus et al., 2015], this oversampling technique was used effectively in other EEG analyses to improve recognition in the dataset.

After the application of the SMOTE method, a random forest classifier was configured and executed. Random forest is an ensemble learning algorithm that uses multiple decision trees as described in detail in [Shiratori et al., 2015].

Several studies have shown that random forest works satisfactorily with features extracted from EEG data with reasonably less training time [Patti et al., 2015; Shiratori et al., 2015; Czarnecki and Gustafsson, 2015]. Additionally, in the seizure classification context, a previous study reached average accuracies greater than 96% using random forest on the epilepsy dataset of the University of Bonn [Wang et al., 2015].

In this study, a random forest classifier was configured to work with 100 decision trees without prune restriction, and with $\log_2(N_{attributes}) + 1$ random features for each tree. We used the random forest implementation of the WEKA API 1.7.0 [Hall et al., 2009].

## 6.4 Results

The method proposed in this work was applied to the physiological signals stored in the CHB-MIT database. Every channel was processed with the STFT transform and PCA for reducing the multivariate set to only one time series. Next, each resultant time series passed through an envelope detector and, finally, the detected envelope was reorganized in blocks of fixed length to be processed by a random forest classifier.

As mentioned previously, two possibilities in the re-sequencing process were considered: Blocks of 30 and 70 seconds. Different sizes showed a influence in the recognition capacity with the transformed signal.

All experiments performed, for each configuration, were assessed using 10-fold cross validation. Regarding the performance indexes, standard parameters were used: Accuracy ($ACC$), specificity ($SPE$), sensitivity ($SEN$), and false positive rate ($FPRe$) measured in samples per hour ($h^{-1}$) [Orosco et al., 2016; Das et al., 2016; Tsiouris et al., 2015; Ahmad et al., 2015; Gill et al., 2015; Zhao et al., 2016; Iqbal et al., 2015]:

$$\text{ACC} = \frac{TN + TP}{TN + TP + FP + FN} \tag{6.5}$$

$$\text{SPE} = \frac{TP}{TP + FN} \tag{6.6}$$

$$\text{SEN} = \frac{TN}{TN + FP} \tag{6.7}$$

$$\text{FPRe} = \frac{FP}{\text{time (in hours)}} \tag{6.8}$$

where $TP$ is the number of true positives, $TN$ is the number of true negatives, $FP$ is the number of false positives, $FN$ is the number of false negatives, and time is a unit of time analysis.

The obtained results, with each execution time $T_{exec}$, were compiled in Table 6.1. It should be noted that the time was calculated using a personal computer with a Core i7 Quad Core processor (3.40 GHz) using Ubuntu 15.10 as operating system.

The experiments showed a good performance with data blocks of 30 seconds, reaching average sensitivity of 89.73%, specificity of 94.77%, $FPRe$ of $6.87h^{-1}$, and accuracy of 92.46%. Nonetheless, increasing the data block size to 70 seconds significantly improved the performance. The average values of the same measures were raised to: 97.12% of sensitivity, 99.29% of specificity, 98.30% of accuracy, and $0.77\,h^{-1}$ of $FPRe$.

## 6.5 Discussion

It should be noted that the database used in this paper is publicly available [Goldberger et al., 2000b]. As a result, there are other seizure detection methods that process the same information to evaluate their algorithms. We selected 9 methods from 7 studies that analyzed the same database to compare with our approach. Their performances are compiled in Table 6.2, although some indexes were not found in their evaluation analyses. For instance, some researches emphasized the importance of the sensibility in the evaluation over other indexes and, thus, presented only sensibility values [Iqbal et al., 2015; Tsiouris et al., 2015].

The current research proposes a different method to process electroencephalographic signals with a high detection accuracy. Comparing with alternative methods, our technique showed the best overall performance. Sensitivity was just behind the technique developed in [Iqbal et al., 2015] that was based on a classification with

**Table 6.1.** Performance evaluation parameters with block sizes of 30 and 70 seconds

| Patient ID | Block size: 30 seconds | | | | | Block size: 70 seconds | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ACC (%) | SEN (%) | SPE (%) | FPRe | $T_{exec}$ (s) | ACC (%) | SEN (%) | SPE (%) | FPRe | $T_{exec}$ (s) |
| chb01 | 96.55 | 94.38 | 98.38 | 2.27 | 60.73 | 99.62 | 99.24 | 99.95 | 0.16 | 87.91 |
| chb02 | 87.46 | 84.40 | 90.05 | 12.13 | 61.45 | 97.26 | 96.38 | 98.00 | 2.42 | 82.6 |
| chb03 | 96.50 | 93.74 | 98.84 | 1.53 | 47.02 | 98.76 | 97.38 | 99.94 | 0.08 | 77.22 |
| chb04 | 94.41 | 91.14 | 97.20 | 15.35 | 369.31 | 99.54 | 99.06 | 99.96 | 0.23 | 509.67 |
| chb05 | 96.25 | 95.61 | 96.79 | 4.33 | 33.21 | 99.82 | 99.73 | 99.91 | 0.13 | 38.45 |
| chb06 | 89.79 | 86.57 | 92.52 | 17.51 | 125.64 | 99.54 | 99.07 | 99.93 | 0.16 | 147.88 |
| chb07 | 88.56 | 85.95 | 90.78 | 21.68 | 137.53 | 99.27 | 98.69 | 99.77 | 0.53 | 185.51 |
| chb08 | 90.89 | 86.17 | 94.85 | 3.55 | 27.75 | 96.41 | 93.50 | 98.85 | 0.78 | 42.49 |
| chb09 | **98.41** | **98.38** | **98.44** | **3.71** | **65.51** | **99.90** | **99.85** | **99.94** | **0.14** | **93.5** |
| chb10 | 94.80 | 91.89 | 97.28 | 4.75 | 75.23 | 98.38 | 97.25 | 99.34 | 1.14 | 115.82 |
| chb11 | 96.39 | 94.21 | 98.24 | 2.12 | 51.01 | 98.81 | 97.83 | 99.63 | 0.44 | 86.89 |
| chb12 | **86.13** | **82.34** | **89.33** | **8.64** | **39.81** | 95.14 | 93.39 | 96.62 | 2.72 | 50.16 |
| chb13 | 92.62 | 90.15 | 94.71 | 6.03 | 55.9 | 97.30 | 95.83 | 98.54 | 1.65 | 78.79 |
| chb14 | 86.73 | 84.26 | 88.83 | 10.03 | 40.29 | 98.08 | 97.16 | 98.86 | 1.02 | 47.16 |
| chb15 | 89.13 | 83.62 | 93.78 | 8.57 | 74.6 | **95.00** | **91.67** | **97.80** | **3.00** | **122.97** |
| chb16 | 87.40 | 83.37 | 90.82 | 6.02 | 30.34 | 97.66 | 95.74 | 99.29 | 0.46 | 35.31 |
| chb17 | 93.84 | 89.55 | 97.47 | 1.75 | 27.36 | 97.77 | 95.34 | 99.83 | 0.12 | 47.74 |
| chb18 | 89.30 | 88.00 | 90.40 | 11.82 | 59.82 | 97.42 | 96.39 | 98.29 | 2.08 | 76.07 |
| chb19 | 96.08 | 97.10 | 95.22 | 4.95 | 35.8 | 99.85 | 99.80 | 99.89 | 0.12 | 37.34 |
| chb20 | 95.76 | 92.70 | 98.36 | 1.56 | 32.47 | 99.21 | 98.40 | 99.90 | 0.09 | 46.39 |
| chb21 | 95.62 | 92.29 | 98.46 | 1.75 | 46.77 | 99.11 | 98.16 | 99.93 | 0.08 | 80.56 |
| chb22 | 89.28 | 84.69 | 93.17 | 7.32 | 50.82 | 98.40 | 96.69 | 99.85 | 0.16 | 84.12 |
| chb23 | 94.28 | 91.19 | 96.90 | 2.88 | 31.32 | 98.25 | 96.43 | 99.78 | 0.20 | 53.77 |
| chb24 | 92.76 | 91.79 | 93.58 | 4.71 | 26.2 | 98.61 | 97.88 | 99.23 | 0.56 | 32.29 |
| **Avg.** | **92.46** | **89.73** | **94.77** | **6.87** | **66.91** | **98.30** | **97.12** | **99.29** | **0.77** | **94.19** |
| **Min.** | 86.13 | 82.34 | 88.83 | 1.53 | 26.20 | 95.00 | 91.67 | 96.62 | 0.08 | 32.29 |
| **Max.** | 98.41 | 98.38 | 98.84 | 21.68 | 369.31 | 99.90 | 99.85 | 99.96 | 3.00 | 509.67 |

entropy-related parameters, which obtained a 100% for $SEN$, but did not present other important evaluation parameters.

Our approach obtained slightly lower values concerning specificity and false positive rate (a difference of 0.70% and 0.44 $h^{-1}$) compared with one of the methods presented in [Orosco et al., 2016]. However, in general, our method demonstrated to be the best approach when all parameter values are taken into account in combination.

A good system response depends on the combination of the internal algorithms selected. Concerning feature selection, the majority of previously proposed techniques rely, at least partially, on spectral features: STFT, wavelet transform, or discrete Fourier transform [Ahmad et al., 2015; Gill et al., 2015; Tsiouris et al., 2015; Das et al., 2016; Orosco et al., 2016]. Furthermore, dimensionality reduction using PCA, or an improved algorithm based on it, proved to lead to learning models that reach accuracies near 100% [Zhao et al., 2016]. The present research combined both processing types using STFT over biosignals and then PCA on the

**Table 6.2.** Performance comparison with other methods

| Study | Year | ACC | SPE | SEN | FPRe | Classifier |
|---|---|---|---|---|---|---|
| [Ahmad et al., 2014] | 2014 | 94.80% | 95.70% | 91.70% | - | Linear SVM |
| [Das et al., 2015] | 2015 | 94.37% | 93.84% | 93.81% | | RBF-SVM |
| [Gill et al., 2015] | 2015 | 86.93% | 86.26% | 87.58% | - | GMM |
| [Iqbal et al., 2015] | 2015 | - | - | 100.00% | 0.8 | - |
| [Tsiouris et al., 2015] | 2015 | - | - | 99.00% | - | - |
| [Orosco et al., 2016] | 2016 | 96.25% | 99.90% | 92.60% | 0.3 | LDA |
| [Orosco et al., 2016] | 2016 | 89.80% | 99.70% | 79.90% | 3.9 | NN |
| [Zhao et al., 2016] | 2016 | 100.00% | - | - | - | SVM |
| The proposed method | 2016 | 98.30% | 99.20% | 97.12% | 0.77 | Random forest |

time-spectrum data. The notable difference with respect to other studies is the use of a singular element typical of amplitude demodulators: An envelope detector to improve the matche between waveforms and seizure marks. Regarding the classification method, several authors used SVM [Ahmad et al., 2015; Das et al., 2016; Zhao et al., 2016], Bayessian classifiers [Gill et al., 2015], linear discriminators or neural networks [Orosco et al., 2016]. In this work, the random forest method demonstrated a very good predictive power, with the advantage of a quick training time [Czarnecki and Gustafsson, 2015].

## 6.6 Conclusions

This study describes a patient-specific system for detecting seizures with a great prediction power. The structure of the proposed approach comprises a sequence of several signal processing and machine learning algorithms. The processing step includes the application of PCA after a STFT transformation with a subsequent envelope shape detection stage. After those procedures, the resultant data are regrouped in fixed length blocks of time that are then given as input to a random forest classifier.

The parameters of the proposed system were optimized to maximize the overall performance. Thus, after training and testing with the CHB-MIT database containing 23 subject cases, we could obtain average values of 98.30% for accuracy, 97.12% for sensitivity, 99.29% for specificity, and $0.77h^{-1}$ as false positive rate. Comparing these indexes with the results of state-of-art alternative approaches, we can conclude that a hardware implementation of our method could lead to a considerable positive impact on epilepsy diagnosis through the automation of seizure detection.

## 6.7 Acknowledgement

This work is supported by grants from Fundação de Amparo á Pesquisa de Minas Gerais (FAPEMIG) and the PAEC agreement between the Organization of American States (OAS) and Universidade Federal de Viçosa (UFV).

# Chapter 7

# Data mining with filter-banks for seizure detection using electroencephalograms

## Abstract

Epilepsy is a very frequent neurological disease, and implies an expensive and specialized diagnosis process based on electroencephalograms and video recordings. We developed a method that requires only the brainwave provided by the difference between two standardly-located electrodes. Our proposed technique separates the original signal using a filter array with three different types of filters, and then extracts several features based on information theory techniques and statistical information. In our study, we found that only 10 features, of which the most importants are related to high frequencies, are required to offer an accuracy of 94%, an specificity of 95%, and a sensitivity of 87% using CART decision trees.

## 7.1 Introduction

Epilepsy is one of the most common and critical neurological diseases. The World Health Organization estimates that between 50 and 200 million people worldwide have some degree of epilepsy [Organization, 2006]. The diagnosis and treatment of this illness use to be an expensive and long process [Tsiouris et al., 2015]. Consequently, in order to reduce time, cost, and specialist-dependency for diagnosis, various alternatives were proposed [Rahman et al., 2015].

The typical method for diagnosis is to analyze the brainwaves or electroencephalograms (EEGs), and use video recordings to determine the patient condition [Tzallas et al., 2009]. Several researches in the field have shown that it is possible to detect seizures using only the brainwaves recorded in the EEGs by means of different techniques that range from analog circuitry to artificial intelligence approaches [Czarnecki and Gustafsson, 2015; Bhavaraju et al., 2006].

EEGs are recordings of the electrical signals of the brain measured in standard points at the scalp. In normal conditions, EEG values oscillate between 5 and 50 microvolts in a bandwidth frequency from 0 to 100 Hz [Vidal et al., 2015]. There are two main issues with brainwaves signal processing: Noise and subject data dependency. The low electrical tension of EEGs make them vulnerable to interference originated in the equipment or even by the patient. EEGs are so sensitive that even a muscular movement could generate some vibrations at the scalp that could lead to misinterpretations [Sanei and Chambers, 2007; Benbadis and Lin, 2008].

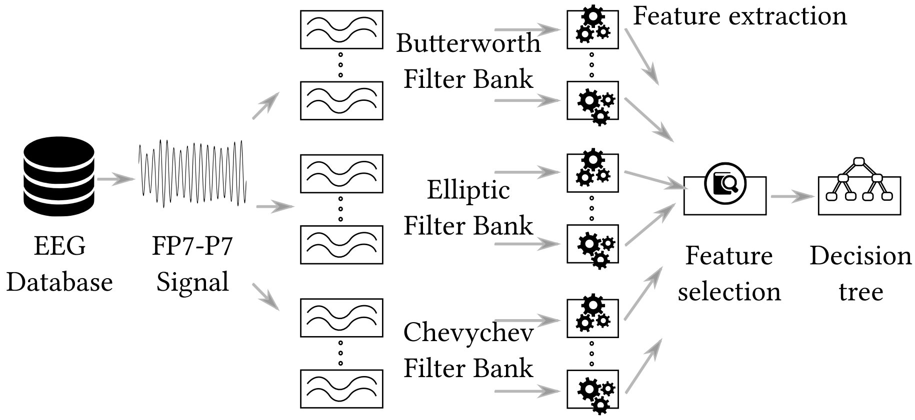
The second usual issue with EEGs is their variations from one patient to another. Some techniques for EEG signal normalization are largely applied such as statistical or spectral normalization. However, according to the objective in the brainwave processing, these methods could reduce the efficiency of the proposed techniques [Shoeb et al., 2004]. For this reason, several algorithms are prepared to work only with one patient at a time. It means that the resultant predictive models generalize poorly, i.e., they are suitable specifically for the patient whose data were used for the training process.

Our method overcomes this limitation relying in two well-known areas of knowledge: Digital signal processing and artificial intelligence. Initially in the text, we describe the basis of our system: The techniques to obtain the features of the original time series, feature selection, and the machine learning algorithm we use. In a second part, we describe in depth the features used to efficiently recognize the seizure instants.

## 7.2 Data source

To perform our experiments, we relied on a trustable and publicly available large dataset: The Children Hospital Boston - Massachusetts Institute of Technology (CHBMIT) dataset from the Physionet project [Goldberger et al., 2000b; Shoeb and Guttag, 2010]. This database stores recordings from 23 different child patients with an average duration of 24 hours for each subject. Each recording was made following

**Figure 7.1.** Workflow of the proposed method




the 10-20 standard layout system for EEG, and was sampled with a frequency of 256Hz.

For the present research, we used only the signal from electrode FP7 (left front-parietal area) and electrode of the position P7 (left parietal lobe) as reference.

## 7.3 Method

We separated our system in three subsystems: Frequency split, feature extraction and analysis, and the detection subsystem itself.

### 7.3.1 Frequency split subsystem

EEG information are saved in the form of frequency variations. For this reason, time-frequency approaches are widely used to process this kind of signals. Low-pass filtering is a classic procedure before signal processing because it removes high frequency noise. Nevertheless, the quality of the output signal depends on the type of filter and its order.

Here, we apply a different methodology: We use a filter bank, which is an array of filters for different frequency intervals, similarly to the idea behind wavelets transform (WT) [Prochazka et al., 2008]. However, in our case, we maintain the filtered signal rather than getting seed codes as in the case of WT. Our proposal is to separate the complete signal in ten segments with a frequency bandwidth of 12.5Hz each. Moreover, there are several types of filters that can be applied with different performance in frequency and phase response. Thus, we selected three different second-order filtering methods: Butterworth (BTW), Chevychev (CHY), and elliptic (ELP) filters.

**Table 7.1.** List of features.

Where $N$ is the number of samples in the window, $x_i$ is a sampled point, $x'_i$ is a normalized sampled point (according to the estimated mean $\mu_x$ and the standard deviation $\sigma_x$), $p_x$ is the probability of a value to appear in the window, and $p'_x$ is the estimated probability of a value to appear in the whole signal.

| **Name** | **Description** | **Definition** |
|---|---|---|
| SignalMean | Mean | $1/N \sum_i x_i$ |
| NormMean | Normalized mean | $1/N \sum_i \frac{x_i - \mu_x}{\sigma_x}$ |
| LProbMean | Local probability mean | $1/N \sum_i p_i$ |
| GProbMean | Global probability mean | $1/N \sum_i p'_i$ |
| SignalInt | Absolute signal integration | $\sum_i \lvert x_i \rvert$ |
| NormInt | Absolute normalized signal integration | $\sum_i \lvert x_i \rvert = \sum_i \left\lvert \frac{x_i - \mu_x}{\sigma_x} \right\rvert$ |
| LProbInt | Local probability integration | $\sum_i \lvert p_i \rvert$ |
| GProbInt | Global probability integration | $\sum_i \lvert p'_i \rvert$ |
| MeanLevel | Mean quantized level | $1/N \sum_i L_i$ |
| MaxLevel | Max quantized level | $\max\{L_i\}_i$ |
| MinLevel | Min quantized level | $\min\{L_i\}_i$ |
| ModeLevels | Mode quantized level | $\text{mode}\{L_i\}_i$ |
| ModeCntLevels | Quantized levels with mode value | $\lVert\{L_i/L_i = \text{mode}\{L_i\}_i\}\rVert$ |
| SignalEntropy | Pseudo-entropy 1 | $1/N \sum_i x_i \log_{10} x_i^2$ |
| NormEntropy | Pseudo-entropy 2 | $1/N \sum_i x'_i \log_{10} x'^2_i$ |
| LProbEntropy | Entropy with local probabilities | $1/N \sum_i p_i \log_{10} p_i^2$ |
| GProbEntropy | Entropy with global probabilities | $1/N \sum_i p'_i \log_{10} p'^2_i$ |

### 7.3.2 Feature analysis subsystem

Thus far, we separated one signal into 30 (10 segments x 3 filters) time series. The next phase is to extract features from these series. One typical approach is to regrouping data in small intervals. In our method, we use a window of one second, or 256 samples, to calculate features.

Two types of features were selected based on statistical calculations and based on information theory measures. For each data slide, we calculate 17 features that are detailed in Table 7.1. Hence, we obtain 510 features in total per second.

**Table 7.2.** Signal transformations.

| Transformation | Definition |
|---|---|
| Linear | $f(x) = x$ |
| Logarithmic | $f(x) = \log(x^2)$ |
| Exponential | $f(x) = \exp(x^{0.5})$ |
| Tangential hyperbolic | $f(x) = \tanh(x^{0.5})$ |

### 7.3.3 Detection subsystem

There are several machine learning algorithms that perform very well with EEG data, such as random forests, and support vector machines. However, they do not provide an easy interpretation of the resultant model. In our study, we prefer to use a CART decision tree as the base algorithm for two reasons. First, it can be easily translated into small programs based on conditionals. Second, and foremost, it can offer a human-understandable interpretation of the data.

As already mentioned, there are 510 features per instance. In order to try to increase the performance of the algorithm, we applied non-linear transformations on the data (Table 7.2). As a result, the number of features was increased to 2040 (510 x 4 transformations). Hence, we tried to reduce this number in order to find the most representatives and offer a better understanding of the process.

To find the minimum quantity of features, we randomly split ten times the dataset in a proportion of 60% for training, and 40% for testing. Next, we used our classification algorithm as a predictor of the efficiency of each column in the dataset, assigning the resultant accuracy and recall as the feature score. After sorting the columns according to the scores, we greedily add each column until the overall performance was not improved. Using this procedure, we found the ten most representative columns.

Testing of our system relied on the software libraries and environments provided by the Weka software (version 3.7.12), Numpy 1.11.1, Scipy 0.17.1, and Scikit-learn 0.17.1.

## 7.4 Results

To validate our proposal, we built ten sets out of the original data, where each such set contained randomly selected instances with a proportion of 60% for training and 40% for testing. In this way, for each experiment with each set, we ensure the response of the system with an average training of approximately 7058.08 seconds

**Table 7.3.** Most representative features.

| Bandwidth frequency [Hz] | Filter type | Feature type | Transformation | Relevance |
|---|---|---|---|---|
| 12.5–25.0 | BTW | LProbInt | Linear | 6th |
| 25.0–37.5 | BTW | LProbInt | Linear | 4th |
| | BTW | LProbMean | Linear | 7th |
| 37.5–50.0 | BTW | ModeLevels | Linear | 2nd |
| | BTW | GProbEntropy | Linear | 8th |
| | BTW | NormMean | Linear | 10th |
| 75.0–87.5 | BTW | MaxLevels | Linear | 9th |
| 87.5–100.0 | ELP | LProbMean | Logarithmic | 3rd |
| | BTW | SignalEntropy | Linear | 5th |
| 112.5–125.0 | CHY | GProbEntropy | Exponential | 1st |

of seizures, and a testing with nearly 4706.36 seconds with seizure events. After performing feature selection, we found 10 representative features that are shown in Table 7.3. Overall, we obtained an accuracy close to 94%, a true positive rate near to 87%, and a specificity of 95%.

There is no study that can be directly compared with our proposal, as the use of signals from only two electrodes was never proposed before. Rubin et al. developed a technique that uses seven electrodes and obtained a general sensitivity of 70%, and a specificity of 96% [Rubin et al., 2014]. Using the CHBMIT dataset, Orosco et al. developed a technique that achieved a specificity greater than 99%, and sensitivity close to 93% [Orosco et al., 2016]. However, Orosco and colleagues used the whole array of electrodes to detect seizures. We consider that our system reflects a positive performance, according to its design constraints, because the use of only two electrodes and the application of a CART decision tree greatly facilitate the construction of small devices to be effectively used by patients.

## 7.5 Discussion

The literature of EEG commonly associates the frequency bands lower than 50Hz to several conditions or activities, whilst the waves with higher frequency are not clearly associated with any particular condition. In our results, we found a surprising fact: the most important feature needed a non-linear transformation, and above all, it was related to the highest analyzed band. Also, from the best 10 features, 4 (40%) were in bands higher than 50Hz, normally ignored in another kind of EEG processing system. Unfortunately, we have no enough information to link any abnormal behavior in these higher frequencies to seizures. However, it can be an important fact to be explored in the future.

Regarding the type of filters, we noted that a second-order filter was sufficient to provide the vast majority of the relevant features of the system. However, the three most important features were obtained from the three kinds of proposed filters. Thus, we cannot ignore them in the overall analysis.

## 7.6 Conclusion

We proposed a new procedure for seizure detection using time series features obtained from filter banks. Our method presents a 4-fold contribution. First, it uses only two electrodes, which allow a practical physical implementation. Second, it provides a very good performance, reaching an accuracy greater than 94%, sensitivity close to 87%, and specificity near to 95%. Third, using decision trees, we are able to translate the predictive tree into programs that can be stored in small devices. Finally, our system was designed to work with independence concerning the origin of the data, i.e., the predictive model built from data of a set of patients might be used for a distinct set of patients, which also facilitates the use of devices in large scale.

## 7.7 Acknowledgment

This work is supported by grants from Fundação de Amparo á Pesquisa de Minas Gerais (FAPEMIG) and the PAEC agreement between the Organization of American States (OAS) and Universidade Federal de Viçosa (UFV).

# Chapter 8

# Conclusions

The objective of this research was to propose an efficient and effective seizure detection method using electroencephalographic signals. In order to comply that goal, we designed two different alternative solutions.

In the first research, we followed the main tendency in the state-of-art studies: Recovering information about the spatially distributed electrical variation of the brainwaves and using spectro-temporal data transformations. Thus, we used all the electrodes available in the dataset and applied a short-time Fourier transform on them. However, we analyzed a new alternative procedure transforming that multivariate signal into a univariate time series. This scheme described a new intermediate graphical representation with a high correlation with seizure intervals. After a specific feature extraction, and a the application of random forest classifier we found, a method that presented a powerful detection performance as could be seen with the typical metrics in the field. As we noted in the corresponding paper, we reached the best overall performance regarding the state-of-art alternatives with a specificity of 99% and a sensitivity of 97% (Figure 8.1.)

Albeit the outstanding results, the main limitation of this proposal is the patient-specific restriction: A suitable model for the patient needs to be built using the own patient data before being applied.

The second research started looking for a method that overcomes the previous constraints. Thus, we used a completely different design. We maintained the concept of spectral and time data, but in a different way: Using filter banks with different types of filters for taking advantage of their characteristics. The proposed method used a decision tree trained with proper linear and nonlinear features. As it was mentioned in the respective chapter, our proposal has a specificity of 95% and a sensitivity of 87% (Figure 8.1.)

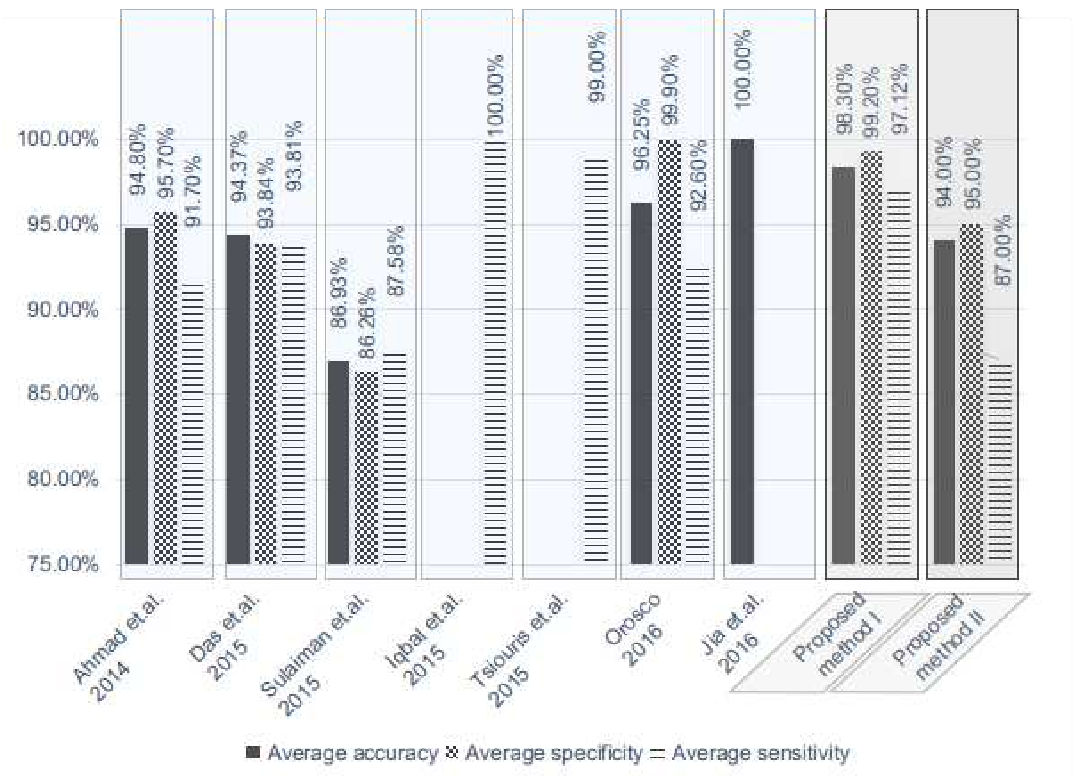



**Figure 8.1.** Performance metrics with our proposed methods compared with the state-of-art

Although these values are not so impressive as seen in previous studies, we should note that the introduced method is not patient-specific, i.e., it could be trained with and applied to data from distinct patients. Also, considering the design conditions of the filter banks and the final decision tree, the system could be implemented easily in a small device.

Both systems, based on distinct approaches, represent solutions for the initial objective with a remarkable performance under different circumstances and conditions.

Many different theories, proposals, and experiments have been left for the future due to limitations of resources and time. In our perspective, the second proposed method offers more challenges and tasks for motivating future studies because of its novelty. The current state of our study give us several open questions: Are there features more relevant than those we used in our methods? Could black-box algorithms, like the random forest or extreme gradient boosting algorithm, improve our performance with small training sets? Is it possible to generalize classification models using small training sets that only consist of a few minutes of recording? Do frequencies higher than 50Hz have an impact in seizure detection? And, probably, the most important question that we were not able to test: Could they be implemented in a device? How will be the performance of these methods in a real environment?

# Bibliography


Aggarwal, V., Jin, W. O., and O'Reilly, U.-M. (2006). Filter approximation using explicit time and frequency domain specifications. page 753. ACM Press.

Ahmad, M. A., Ayaz, Y., Jamil, M., Omer Gillani, S., Rasheed, M. B., Imran, M., Khan, N. A., Majeed, W., and Javaid, N. (2015). Comparative Analysis of Classifiers for Developing an Adaptive Computer-Assisted EEG Analysis System for Diagnosing Epilepsy. *BioMed Research International*, 2015:1--14.

Ahmad, M. A., Khan, N. A., and Majeed, W. (2014). Computer Assisted Analysis System of Electroencephalogram for Diagnosing Epilepsy. pages 3386--3391. IEEE.

Akalin Acar, Z. and Makeig, S. (2013). Effects of Forward Model Errors on EEG Source Localization. *Brain Topography*, 26:378--396.

Al-Fahoum, A. S. and Al-Fraihat, A. A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neuroscience*, 2014:1--7.

Alotaiby, T. N., Abd El-Samie, F. E., Alshebeili, S. A., Aljibreen, K. H., and Alkhanen, E. (2015). Seizure detection with common spatial pattern and Support Vector Machines. In *Information and Communication Technology Research (ICTRC), 2015 International Conference on*, pages 152--155. IEEE.

Alotaiby, T. N., Alshebeili, S. A., Alshawi, T., Ahmad, I., and El-Samie, F. E. A. (2014). EEG seizure detection and prediction algorithms: a survey. *EURASIP Journal on Advances in Signal Processing*, 2014:1--21.

Awan, S. E., Akram, M. U., and Khan, M. A. (2015). Analysis of EEG Signals for Detection of Epileptic Seizure.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Filho, W. J., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. 513:532--541.

Barkley, G. L. (2004). Controversies in neurophysiology. MEG is superior to EEG in localization of interictal epileptiform activity: Pro. 115:1001--1009.

Benbadis, S. R. (2006). Introduction to sleep electroencephalography. *Sleep: A Comprehensive Handbook. USA: John Wiley & Sons*, pages 989--1024.

Benbadis, S. R. and Lin, K. (2008). Errors in EEG interpretation and misdiagnosis of epilepsy. 59:267--271.

Berger, H. (1929). Uber das elektrenkephalogramm des menschen. 87:527--570.

Bhagwat, S. D. and Jain, V. (2013). EEG Data Sets Signal Processing Using Wavelet Transforms. *no*, 6:108--111.

Bhavaraju, N., Frei, M., and Osorio, I. (2006). Analog Seizure Detection and Performance Evaluation. *IEEE Transactions on Biomedical Engineering*, 53:238--245.

Birjandtalab, J., Pouyan, M. B., and Nourani, M. (2016). Nonlinear dimension reduction for EEG-based epileptic seizure detection. pages 595--598. IEEE.

Bömers, F. (2000). Wavelets in real time digital audio processing: Analysis and sample implementations. *University of Manheim, Dept. of Computer Science IV, Master's thesis.*

Breiman, L. (2001). Random forests. *Machine learning*, 45:5--32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees. wadsworth & brooks. *Monterey, CA*.

Chan, J. C.-W. and Paelinckx, D. (2008). Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112:2999--3011.

Czarnecki, M. and Gustafsson, N. (2015). *Machine learning for detection of epileptic seizures.* PhD thesis, Chalmers University of Technology, Gothenburg, Sweden.

da Silva, F. L. (2009). EEG: Origin and Measurement. In Mulert, C. and Lemieux, L., editors, *EEG - fMRI*, pages 19--38. Springer Berlin Heidelberg, Berlin, Heidelberg.

Das, A. B., Bhuiyan, M. I. H., and Alam, S. M. S. (2016). Classification of EEG signals using normal inverse Gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection. *Signal, Image and Video Processing*, 10:259--266.

Das, A. B., Pantho, M. J. H., and Bhuiyan, M. I. H. (2015). Discrimination of scalp EEG signals in wavelet transform domain and channel selection for the patient-invariant seizure detection. pages 77--80. IEEE.

Daubechies, I. (1992). *Ten lectures on wavelets*. CBMS-NSF regional conference series in applied mathematics. Society for Industrial and Applied Mathematics.

De Bock, K. W., Coussement, K., and Van den Poel, D. (2010). Ensemble classification based on generalized additive models. *Computational Statistics & Data Analysis*, 54:1535--1546.

Deng, H., Runger, G., Tuv, E., and Vladimir, M. (2013). A time series forest for classification and feature extraction. *Information Sciences*, 239:142--153.

Dhankhar, P. and Khaleri, S. (2014). Eye Blink Artifact Removal in EEG Using Adaptive Fir Filter-A Review. *International Journal of Emerging Technology and Advanced Engineering*, 4:700--703.

Dimopoulos, H. G. (2007). Optimal Use of Some Classical Approximations in Filter Design. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 54:780--784.

Djemili, R., Bourouba, H., and Amara Korba, M. (2016). Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals. *Biocybernetics and Biomedical Engineering*, 36:285--291.

Dong, E. and Liang, Z. (2014). The multi-frequency EEG rhythms modeling based on two-parameter bifurcation of neural mass model. pages 1564--1569. IEEE.

Duo Chen, Suiren Wan, and Forrest Sheng Bao (2015). EEG-based seizure detection using discrete wavelet transform through full-level decomposition. pages 1596--1602. IEEE.

Durka, P. (2007). *Matching pursuit and unification in EEG analysis*. Artech House engineering in medicine and biology series. Artech House.

Fatichah, C., Iliyasu, A. M., Abuhasel, K. A., Suciati, N., and Al-Qodah, M. A. (2014). Principal component analysis-based neural network with fuzzy membership function for epileptic seizure detection. pages 186--191. IEEE.

Fergus, P., Hignett, D., Hussain, A., Al-Jumeily, D., and Abdel-Aziz, K. (2015). Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques. *BioMed Research International*, 2015:1--17.

Freeman, W. J. and Quiroga, R. Q. (2013). *Imaging Brain Function With EEG*. Springer New York.

Gill, A. F., Fatima, S. A., Usman Akram, M., Khawaja, S. G., and Awan, S. E. (2015). Analysis of EEG Signals for Detection of Epileptic Seizure Using Hybrid Feature Set. In Sulaiman, H. A., Othman, M. A., Abd. Aziz, M. Z. A., and Abd Malek, M. F., editors, *Theory and Applications of Applied Electromagnetics*, volume 344, pages 49--57. Springer International Publishing, Cham.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000a). Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101:e215--e220.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000b). PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals. 101:e215--e220.

Guo, L., Rivero, D., Dorado, J., Rabunal, J. R., and Pazos, A. (2010). Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods*, 191:101--109.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11:10--18.

Ham, J., Chen, Y., Crawford, M. M., and Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43:492--501.

Hesse, W., Moller, E., Arnold, M., and Schack, B. (2003). The use of time-variant EEG granger causality for inspecting directed interdependencies of neural assemblies. 124:27--44.

Hu, B., Liu, Q., Zhao, Q., Qi, Y., and Peng, H. (2011). A Real-Time Electroencephalogram (EEG) Based Individual Identification Interface for Mobile Security in Ubiquitous Environment. pages 436--441. IEEE.

Iqbal, S., Khan, Y. U., and Farooq, O. (2015). Nonlinear analysis of EEG for seizure prediction. pages 1--5. IEEE.

Jovic, A., Suc, L., and Bogunovic, N. (2013). Feature extraction from electroencephalographic records using EEGFrame framework. In *Information & Communication Technology Electronics & Microelectronics (MIPRO), 2013 36th International Convention on*, pages 965--970. IEEE.

Kamel, N. and Malik, A. S., editors (2015). *EEG/ERP analysis: methods and applications*. CRC Press, Taylor & Francis Group, Boca Raton.

Kiasaleh, K. (2015). *Biological Signals Classification and Analysis*. Lecture Notes in Bioengineering. Springer Berlin Heidelberg.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., Isenhart, R., and John, E. (2002). Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *NeuroImage*, 16:41--48.

Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26:159--190.

Laghari, W. M., Baloch, M. U., Mengal, M. A., and Shah, S. J. (2014). Performance Analysis of Analog Butterworth Low Pass Filter as Compared to Chebyshev Type-I Filter, Chebyshev Type-II Filter and Elliptical Filter. *Circuits and Systems*, 05:209--216.

Langley, P. and Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38:54--64.

Lawrence, R. L. and Wright, A. (2001). Rule-based classification systems using classification and regression tree (cart) analysis. *Photogrammetric engineering and remote sensing*, 67:1137--1142.

Lee, G., Kwon, M., Kavuri Sri, S., and Lee, M. (2014). Emotion recognition based on 3d fuzzy visual and EEG features in movie clips. *Neurocomputing*, 144:560--568.

Leuthardt, E., Miller, K., Schalk, G., Rao, R., and Ojemann, J. (2006). Electrocorticography-based brain computer interface–the seattle experience. 14:194--198.

Lewis, R. J. (2000). An introduction to classification and regression tree (cart) analysis. In *Annual meeting of the society for academic emergency medicine in San Francisco, California*, pages 1--14.

Liang, H., Bronzino, J. D., and Peterson, D. R., editors (2013). *Biosignal processing: principles and practices*. CRC Press/Taylor & Francis, Boca Raton.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1:14--23.

Malutan, R., Gomez Vilda, P., and Borda, M. (2013). A hybrid unsupervised and supervised clustering applied to microarray data. *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, 2.

Mandal, M. K. and Asif, A. (2007). *Continuous and discrete time signals and systems*. Cambridge University Press, Cambridge, UK ; New York. OCLC: ocm74523222.

Martinez, E. and Martin, M. T. (2011). Tecnicas de clasificacion de opiniones aplicadas a un corpus en espanol. *Procesamiento del Lenguaje Natural*, 47:163--170.

Melville, J. L., Burke, E. K., and Hirst, J. D. (2009). Machine learning in virtual screening. *Combinatorial chemistry & high throughput screening*, 12:332--343.

Miller, R. (2007). Theory of the normal waking EEG: From single neurones to waveforms in the alpha, beta and gamma frequency ranges. *International Journal of Psychophysiology*, 64:18--23.

Mitchell, T. M. (2006). *The discipline of machine learning*, volume 9. Carnegie Mellon University, School of Computer Science, Machine Learning Department.

Napflin, M., Wildi, M., and Sarnthein, J. (2007). Test-retest reliability of resting EEG spectra validates a statistical signature of persons. *Clinical Neurophysiology*, 118:2519--2524.

Natarajan, K., Acharya, R., Alias, F., Tiboleng, T., and Puthusserypady, S. K. (2004). Nonlinear analysis of EEG signals at different mental states. *BioMedical Engineering OnLine*, 3:7.

Nidal, K. and Malik, A. S. (2014). *EEG/ERP Analysis: Methods and Applications*. CRC Press.

Oppenheim, A. V., Willsky, A. S., Hamid Nawab, S., Mata Hernandez, G., and Suarez Fernandez, A. (1998). *Senales y sistemas*. Prentice Hall, Mexico [etc. OCLC: 626650148.

Organization, W. H., editor (2006). *Neurological disorders: public health challenges*. World Health Organization, Geneva. OCLC: ocn122438264.

Orosco, L., Correa, A. G., Diez, P., and Laciar, E. (2016). Patient non-specific algorithm for seizures detection in scalp EEG. *Computers in Biology and Medicine*, 71:128--134.

Orosco, L., Correa, A. G., and Laciar, E. (2013). Review: a survey of performance and techniques for automatic epilepsy detection. *Journal of Medical and Biological Engineering*, 33:526--537.

Page, A., Pramod Tim Oates, S., and Mohsenin, T. (2015). An ultra low power feature extraction and classification system for wearable seizure detection. pages 7111--7114. IEEE.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26:217--222.

Palaniappan, R. (2006). Electroencephalogram signals from imagined activities: a novel biometric identifier for a small population. In *Intelligent Data Engineering and Automated Learning-IDEAL 2006*, pages 604--611. Springer.

Patel, R., Sengottuvel, S., Janawadkar, M., Gireesan, K., Radhakrishnan, T., and Mariyappa, N. (2015). Ocular artifact suppression from EEG using ensemble empirical mode decomposition with principal component analysis. *Computers & Electrical Engineering*.

Patti, C. R., Shahrbabaki, S. S., Dissanayaka, C., and Cvetkovic, D. (2015). Application of random forest classifier for automatic sleep spindle detection. In *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*, pages 1--4. IEEE.

Peker, M., Arslan, A., Sen, B., Çelebi, F. V., and But, A. (2015). A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF). In *Innovations in Intelligent SysTems and Applications (INISTA), 2015 International Symposium on*, pages 1--8. IEEE.

Podder, P., Mehedi Hasan, M., Rafiqul Islam, M., and Sayeed, M. (2014). Design and Implementation of Butterworth, Chebyshev-I and Elliptic Filter for Speech Signal Analysis. *International Journal of Computer Applications*, 98:12--18.

Prochazka, A., Kukal, J., and Vysata, O. (2008). Wavelet transform use for feature extraction and EEG signal segments classification. In *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, pages 719--722. IEEE.

Rahman, F. A., Othman, M. F., and Shaharuddin, N. A. (2015). Analysis Methods of EEG Signals: A Review in EEG Application for Brain Disorder. *Jurnal Teknologi*, 72.

Rahul Kumar Chaurasiya, Jain, K., Goutam, S., and Manisha (2015). Epileptic seizure detection using HHT and SVM. pages 1--6. IEEE.

Razi, M. A. and Athappilly, K. (2005). A comparative predictive analysis of neural networks (nns), nonlinear regression and classification and regression tree (cart) models. *Expert Systems with Applications*, 29:65--74.

Rehman, A., Mustafa, M., Israr, I., and Yaqoob, M. (2013). Survey of wearable sensors with comparative study of noise reduction ecg filters. *Int. J. Com. Net. Tech*, 1:61--81.

Roget, L. B. J., Hernandez, R. F., and Plasencia, C. P. (2016). Conjunto de herramientas citogeneticas para el trabajo con cromosomas.

Rubin, M. N., Jeffery, O. J., Fugate, J. E., Britton, J. W., Cascino, G. D., Worrell, G. A., Hocker, S. E., Wijdicks, E. F., and Rabinstein, A. A. (2014). Efficacy of a reduced electroencephalography electrode array for detection of seizures. *The Neurohospitalist*, 4:6--8.

Sahu, O., Anand, V., Kanhangad, V., and Pachori, R. B. (2015). Classification of magnetic resonance brain images using bi-dimensional empirical mode decomposition and autoregressive model. *Biomedical Engineering Letters*, 5:311--320.

Sakkalis, V., editor (2015). *Modern electroencephalographic assessment techniques: theory and applications*. Neuromethods. Humana Press, New York. OCLC: ocn892869727.

Samiee, K., Kovacs, P., and Gabbouj, M. (2015). Epileptic Seizure Classification of EEG Time-Series Using Rational Discrete Short-Time Fourier Transform. *IEEE Transactions on Biomedical Engineering*, 62:541--552.

Sanei, S. (2013). *Adaptive processing of brain signals*. John Wiley & Sons Inc.

Sanei, S. and Chambers, J. (2007). *EEG signal processing*. John Wiley & Sons.

Saranya, S., Loganathan, V., and RamaPraba, P. S. (2015). Efficient feature extraction and classification of chromosomes. pages 1--7. IEEE.

Shiratori, T., Tsubakida, H., Ishiyama, A., and Ono, Y. (2015). Three-class classification of motor imagery EEG data including "rest state" using filter-bank multi-class Common Spatial pattern. In *Brain-Computer Interface (BCI), 2015 3rd International Winter Conference on*, pages 1--4. IEEE.

Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Treves, S. T., and Guttag, J. (2004). Patient-specific seizure onset detection. *Epilepsy & Behavior*, 5:483--498.

Shoeb, A. H. and Guttag, J. V. (2010). Application of machine learning to epileptic seizure detection. pages 975--982.

Sierra-Marcos, A., Scheuer, M. L., and Rossetti, A. O. (2015). Seizure detection with automated EEG analysis: A validation study focusing on periodic patterns. *Clinical Neurophysiology*, 126:456--462.

Speybroeck, N. (2012). Classification and regression trees. *International journal of public health*, 57:243--246.

Spyrou, L., Martin-Lopez, D., Valentin, A., Alarcon, G., and Sanei, S. (2016). Detection of Intracranial Signatures of Interictal Epileptiform Discharges from Concurrent Scalp EEG. *International Journal of Neural Systems*, 26:1650016.

Svetnik, V., Liaw, A., Tong, C., and Wang, T. (2004). Application of breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In *International Workshop on Multiple Classifier Systems*, pages 334--343. Springer.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement science review*, 2:1--11.

Tsiouris, K. M., Konitsiotis, S., Markoula, S., Koutsouris, D. D., Sakellarios, A. I., and Fotiadis, D. I. (2015). An unsupervised methodology for the detection of epileptic seizures in long-term EEG signals. pages 1--4. IEEE.

Tu, W. and Sun, S. (2012). A subject transfer framework for EEG classification. *Neurocomputing*, 82:109--116.

Tulchinsky, T. H., Varavikova, E., and Bickford, J. D. (2014). *The new public health*. Elsevier/AP, Amsterdam ; Boston, third edition edition.

Tzallas, A., Tsipouras, M., and Fotiadis, D. (2009). Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis. *IEEE Transactions on Information Technology in Biomedicine*, 13:703--710.

Urcuqui, C. and Navarro, A. (2016). Clasificadores de machine learning para analisis de malware en Android. In *IEEE 11 Congreso Colombiano de Computacion*.

Verleysen, M., Universite catholique de Louvain, and Katholieke Universiteit Leuven, editors (2014). *Proceedings / 22nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2014: Bruges, Belgium, April 23 - 25, 2014*. Ciaco, Louvain-la-Neuve.

Vidal, F., Burle, B., Spieser, L., Carbonnell, L., Meckler, C., Casini, L., and Hasbroucq, T. (2015). Linking EEG signals, brain functions and mental operations: Advantages of the Laplacian transformation. *International Journal of Psychophysiology*, 97:221--232.

Vogel, F. (1970). The genetic basis of the normal human electroencephalogram (EEG). 10:91--114.

Wang, G., Deng, Z., and Choi, K.-S. (2015). Detection of Epileptic Seizures in EEG Signals with Rule-Based Interpretation by Random Forest Approach. In Huang, D.-S. and Han, K., editors, *Advanced Intelligent Computing Theories and Applications*, volume 9227, pages 738--744. Springer International Publishing, Cham.

WHO, W. H. O., editor (2006). *Neurological disorders: public health challenges*. World Health Organization, Geneva. OCLC: ocn122438264.

Widmann, A., Schroger, E., and Maess, B. (2015). Digital filter design for electrophysiological data - a practical approach. 250:34--46.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural computation*, 8:1341--1390.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14:1--37.

Yan, A., Zhou, W., Yuan, Q., Yuan, S., Wu, Q., Zhao, X., and Wang, J. (2015). Automatic seizure detection using Stockwell transform and boosting algorithm for long-term EEG. *Epilepsy & Behavior*, 45:8--14.

Zhang, Z. and Smith, R. L. (2001). Modeling financial time series data as moving maxima processes.

Zhao, Q., Peng, H., Hu, B., Liu, Q., Liu, L., Qi, Y., and Li, L. (2010). Improving individual identification in security check with an EEG based biometric solution. In *Brain Informatics*, pages 145--155. Springer.

Zhao, W., Qu, J., Chai, Y., and Tang, J. (2016). Classification of Seizure in EEG Signals Based on KPCA and SVM. In Jia, Y., Du, J., Li, H., and Zhang, W., editors, *Proceedings of the 2015 Chinese Intelligent Systems Conference*, pages 201--207. Springer Berlin Heidelberg, Berlin, Heidelberg.