Analysis for Automatic Detection of Epileptic Seizure from EEG signals

Giulia Pellegrino
Matricola: 1056238
Relatore: Professor Gianna Maria Toffolo
Correlatori: Nicola Rizzo e Gianpietro Favaro
Università degli Studi di Padova
Facoltà di Ingegneria
Laurea Specialistica in Bioingegneria

14 Luglio 2014

There is a driving force more
powerful than steam, electricity
and atomic energy: the will.

Albert Einstein

Contents

Chapter 1

Introduction

Epilepsy (from the Ancient Greek meaning “to seize, possess, or afflict”) is a group of long-term neurological disorders characterized by epileptic seizures.

Epileptic seizures are neurological dysfunctions that are manifested in abnormal electrical activity of the brain. Behavioural correlates, such as convulsions, are sometimes associated with seizures. There are, however, seizures that do not have clear external manifestations. These non-convulsive seizures can be detected only by monitoring brain activity.
One way to investigate the electrical activity of the brain is to record scalp potential resulting from brain activity. The recorded signal, i.e., potential difference between two positions, is called electroencephalogram.
Long-term measurements generate a lot of data and manually reviewing all of it is an exhausting task. There is a clear need for an automatic seizure detection method.

The term *seizure detection* generally refers to the use of an automated algorithm (a *seizure detection algorithm* or *SDA*) to recognize that a seizure is occurring (or has occurred) through analysis of biologic signals recorded from a patient with epilepsy.
During seizure, there is abnormal electrical activity in the brain, that is reflected on scalp potentials and hence can be recorded with EEG.
Essentially, the goal is to receive and analyze a set of signals and transform the information they contain into an output signal or indicator of whether or not the patient is in a state of seizure.
Important objectives are to perform this transformation as quickly, efficiently and as accurately possible.

The present work was developed in cooperation with Micromed S.p.a., an Italian company manufacturing electro medical devices for Neurophysiology Diagnostics.
It is currently one of the few company able to develop, design and manufacture in house both hardware and software of the systems.

Goal

The aim of this project is to evaluate two seizure detection algorithm and to compare their performance using validation criteria of performance, starting from exhaustive state of the art literature study.

There are numerous seizure detection algorithms described in the literature. The first attempts at automated seizure detection were made by Gotman who tried to quantify EEG transients/spikes (Gotman and Gloor 1976, Ives et al. 1976) and by Lopes Da Silva about nonstationarities (1975).

In 1976, **Gotman and Gloor** [7] proposed a method of recognition and quantification of interictal epileptic activity (sharp waves and spikes) in human scalp EEG.
In order to develop a method for automatic recognition, the EEG of each channel was broken down into half waves. A wave was characterized by the durations and amplitudes of its two component half waves, by the second derivative at its apex measured relative to the background activity, and by the duration and amplitude of the following half wave. This method gave the basis to the work in the field of seizure detection. The main limitation of the method was the absence of a solid definition for an interictal epileptic event.
In 1982, **Gotman** [8] proposed an improved method for automatic detection of seizures in EEG and after this, many methods have been proposed to detect the seizures.

Qu and Gotman [15] proposed a patient specific seizure onset detection method and achieved a sensitivity of 100% with mean latency of 9.4 seconds. The false positive rate declared were 0.02 per hour but the drawback of this method was the need of template for the detection of seizures.

In 2004, **Gotman and Saab** [17] designed an onset detection system. It was tested using scalp EEG and reported sensitivity of 77.9%, false detection rate of 0.9 per hour and median detection delay of 9.8 seconds.

Sorensen et al (2010) [22] used matching pursuit algorithm and achieved 78-100% sensitivity with 5-18 seconds delay in seizure onset detection while at the same time 0.2- 5.3 false positives per hour were reported. The method was evaluated using both scalp and intracranial EEG.
Also in 2010 **Shoeb and Guttag** [21] reported 96% sensitivity and mean detection delay of 4.6 seconds working on free online database CHB-MIT with the application of Machine Learning to epileptic seizure detection.

Kharbouch et al [11] proposed a method for seizure detection from iEEG. The data of 10 patients was utilized to extract both temporal and spectral features. The method detected 97% of 67 test seizures with a median detection delay of 5 seconds and a median false detection rate of 0.6 per 24 hour.

Commonly used measures for EEG quantification include:
(i) amplitude and/or signal power, often restricted to a particular frequency band using a filter,
(ii) frequency and/or duration changes in the signal,
(iii) phase variable changes,
(iv) rhythmicity changes, and
(v) a measure of distance between the signal segment and a template signal with known morphology.

These quantities may also be combined to derive other measures, such as measures of similarity between the power spectral densities obtained from different signal epochs (Murro et al. 1991, Alarcon et al. 1995, Gabor and Seyal 1996).

The most widely used systems are those developed in Stellate and Persyst companies.
Gotman, involved in the field of EEG monitoring and seizure detection since the 1970s, formed in 1986 Stellate Systems and his algorithms, that we will see below, are distributed by this company.

Persyst Development Corporation also offers seizure detection software. The algorithm, Reveal Rosetta, is also promised to hold potential for ICU use. The structure of the algorithm is largely unpublished. In [26] the seizure detection Reveal algorithm is described in order to evaluate and compare with two other algorithms.

This study proposes a method of automatic detection of epileptic seizure us-

ing wavelet based features, in which normal and epileptic EEG signals were classified using neural network.
This algorithm is compared to a parametric algorithm based on Autoregressive (AR) model parameters.

Summary

In *Chapter 2*, a brief introduction to the problem is given, and there is a theory introduction to the project, where EEG and Epilepsy are described in order to give to the reader the basics required to comprehend the study.

The first step involved a review of state of the art literature, and in particular some studies described in *Chapter 3* were taken as reference for the design of the detection algorithm.

Chapter 4 presents the data acquisition system, the database used for this this study is described and commented and there is a description of data preprocessing.

From *Chapter 5* to *Chapter 7* there is a description of the chosen methods for the design and the validation of a classifier, in the former the two method for feature extraction are presented: one based on Wavelet Transform and one based on AR modeling.
The other two chapters outlines the proposed classification method, based on Artificial Neural Network, and the validation method.

Chapter 8 describes in detail the implementation of the algorithm and the reason that led me to make decision about the chosen parameters.

Chapter 9 and *Chapter 10* discusses the results of the seizure detection algorithms, summarizes the conclusions and provides a discussion about the research methods.
Possible ways to improve the results and algorithms are given as possible future work.

Chapter 2

Epilepsy and EEG

Epilepsy is an important neurological disorder, characterized by recurrent seizures. In the world epilepsy affects an estimated 43 million people. Thirty percent of epilepsy patients between the ages of 5 and 25 develop seizures related to illness or accidents involving an injury to the head. As many as 50 percent of epilepsies continue into adulthood.
Epilepsy is classified as the second most serious neurological condition known to man, after stroke. It affects nearly 50 million people around the world, which is approximately 1% of the world population.

In this chapter a general introduction to epilepsy, epileptic seizures and EEG are given.

2.1 Epilepsy

Giving a specific definition of Epilepsy is not easy because the disease cannot be identified as one condition but rather as a syndrome, where many different symptoms are present. In 2005 a definition of epilepsy and epileptic seizure was given, when the *International League Against Epilepsy* (ILAE) and the *International Bureau for Epilepsy* (IBE) agreed on these terms:

Epileptic seizure: transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain [6].

Epilepsy: disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure [6].

It is important to distinguish Epilepsy and Epileptic Seizures because the last one is a phenomena, instead the Epilepsy is a chronic pathology that consists in seizures. There are over 40 different types of epilepsy with own definite syntomps, characterized by their own specific combination of seizure type, EEG characteristics, typical age of onset and treatment.

In the following section the seizure types are classified in an internationally recognized scheme and the type of epilepsies related to absence seizures are described in the subsequent section.

2.1.1 Seizure Types

Types of seizures, their intensity, frequency and duration vary a lot between patients, but is common that the same pattern is repeated on each occasion on a given patient.
A classification scheme for seizure classification proposed in 1981 by ILAE [1] was shown in Figure 2.1. Classifying the type of seizure helps doctors diagnose whether or not a patient has epilepsy.
Based on the type of behavior and brain activity, the first main division is between generalized seizures, which affects the whole brain, and the partial seizures, which affects only a part of the brain.

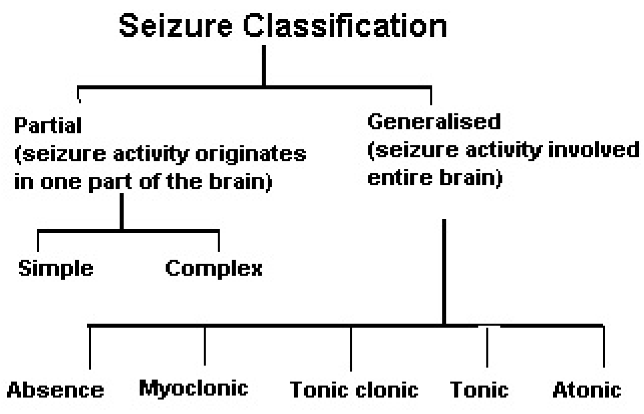


Figure 2.1: Seizure Classification according to ILAE [1]

2.1.2 Absence Seizures

This type of seizures were first observed by Poupart in 1705 and defined as *petite access* by Tissot in 1770. The term *absence* was coined and used for the first time by Calmell in 1824 [24]. The most common type of pediatric seizures is characterized by 3-Hz spike-and-slow-wave complexes in the EEG (Figure 2.2), an abrupt onset and ending, duration usually less of 30 seconds and loss of awareness, as shown in table 2.1.

Absence Seizure Feature	
Onset	**Abrupt**
Duration	**Usually $<$ 30s**
Automatisms	**Duration dependent**
Awareness	**No**
Ending	**Abrupt**

Table 2.1: Features of absence seizures

They are often characterized by slow spike-wave complexes of 1.5-2.5 Hz, which are also known as sharp-and-slow-wave complexes. Seizure of this type are referred to as atypical absence seizures.

2.2 EEG

Human EEG was first recorded by Hans Berger in the 1920s. As a pioneer in the field, it was he who coined the term electroencephalogram. His work was based on initial animal brain function studies performed by Richard Caton in the 19th century.

In the field of seizure detection, the earliest studies were performed in the 1930s when Fisher and Lowenback described epileptiform spikes. Throughout the latter half of the 20th century, with the dawn of digital recording techniques and widely accessible computing power, we have seen a surge of studies that describe both the origins of EEG and what clinicians and researchers can infer from it. Today, EEG is widely accepted as a standard measurement technique.

A modern EEG device consists of a set of electrodes, an amplifier, a data storage unit, and a display unit.

It is known that epilepsy can be detected based on the electroencephalogram (EEG) signal analysis.
EEG signal of epilepsy patients during a seizure shows patterns which are significantly different compared to the normal state of the brain with respect to space, time and frequency patterns. In recent years, many research efforts demonstrated the feasibility of using intracranial or scalp EEG signal to predict and detect seizures.

The electroencephalogram is obtained by recording the activity of the cerebral cortex, by applying the electrodes to the head. The electrodes, suitably positioned on the scalp, record events of electrical nature that occur in the underlying cortex.
The EEG is a mean of investigation in brain diseases mainly thanks to its high temporal resolution allowing to detect in real time the events taking place in the scale of milliseconds. It is used in cases of seizure disorders such as epilepsy or to report the presence of brain tumors.

The EEG detects the potential difference between an active electrode, placed above the seat, the place of the neural activity, and an indifferent electrode, placed at a certain distance from the first. It greatly depends on how synchronized the activity of neurons involved.

In general, in normal man, the electrical potentials recorded from the scalp variys between 20 and 100 microvolts and are characterized by frequencies that ranging from 1 to about 60 Hz. Approximately 90% of the EEG signal seems highly random in the sense that it is not attributable to particular states of mind and being difficult to interpret, usually called "background activity" or, more "noise".
Apart from this background noise, in the course of numerous studies of electroencephalography particular waveforms were identified, said oscillations or EEG rhythms, which may be good indicators of disease or injury, or simply the state of relaxation of the subject.
These are often related to specific behavioral states (such as levels of activation, sleep and wakefulness) and pathological (access or coma). The rhythms are classified on the basis of the frequency range within which vary, and each interval is identified by a Greek letter.

2.2.1 main EEG rhythms

- **Delta rhythms** (δ) are rather slow , less than 4 Hz , often of great size , and are a characteristic feature of deep sleep, in pathological conditions such as coma or cancers.
- **Theta waves** (θ) are of 3-7 Hz and recorded during some sleep states (have the greatest amplitude).
- **Apha** (α) is approximately 8-13 Hz and are associated with relaxed states of wakefulness (best record from the occipital and parietal lobes, the back of the brain). The desynchronization of the alpha rhythm, i.e. the decrease in the amplitude of the waves, could be related to increased availability of networks or cortical sensory input to the motor command.
- **Beta** (β) **e Gamma rhythms** (γ) consist of all frequencies greater than about 14 Hz (14-30 Hz and > 30 Hz , respectively) and are indicative of a bark activated (normally observed at the level of the frontal areas, but you can also record from other cortical regions; during intense mental activity have the minimum width).

ABSENCE: It is characterized by multiple typical absence seizures, accompanied with bilateral, symmetrical, and synchronous discharges of 3-Hz generalized spike and waves on electroencephalography (EEG).

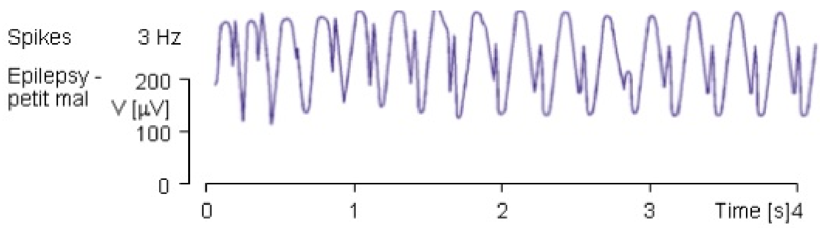


Figure 2.2: The spike and slow wave complexes in a patient with absence seizures, from [13]

2.2.2 Disposition of the electrodes: the international system 10/20

In order to achieve a proper recording of EEG signals is essential the correct positioning of the electrodes on the scalp and the correct derivation of the signal by the same. The reference system currently accepted, namely, that

allows a relative consistency and comparability between different exams, is the international system 10/20. The standard positions are identified by a letter and a number. The letters are:
F = frontal
T = temporal
C = central
P = parietal
O = occipital
A = ear

The even numbers relate to the right hemisphere, the odd to the left. Z indicates the center line. A1 and A2 are at the lobes of the ears and are used as a reference.

- Placement Anterior-posterior: starting from the top you measure the distance between the nasion and inion and the distances are calculated as shown in figure 2.3.

- Placement in the coronal plane: measuring the distance between the right and left auricular points and the distances are calculated as shown in Figure 2.3.

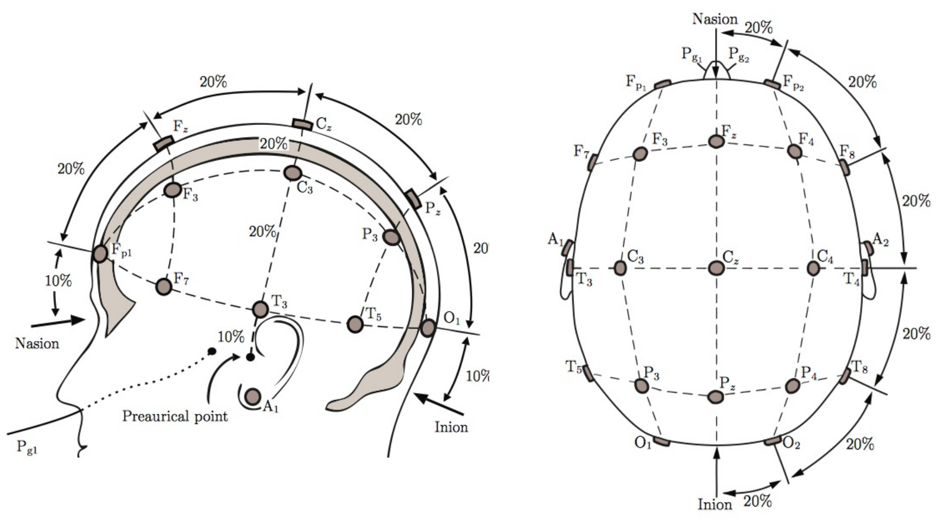


Figure 2.3: EEG modello 10-20

An EEG channel, or derivation, at its simplest form consists of two electrodes that are connected to an amplifier. The output signal is the amplified potential difference between the electrodes. In the common reference mode, one

electrode is used as a reference for all other electrodes. A ground electrode can be used to reduce mains interference. A collection of derivations is called a montage.
The standard montage is 19 electrodes. It is possible to increase the spatial information provided by EEG using a greater number of electrodes. In this case we speak of the international extended system 10/20, and the electrodes in the intermediate positions are called supernumerary electrodes. In Figure 2.4, the position of the 19 standard electrodes is shown in red, the position of the electrodes in additional green (32-electrode assembly) and white (64-electrode assembly).

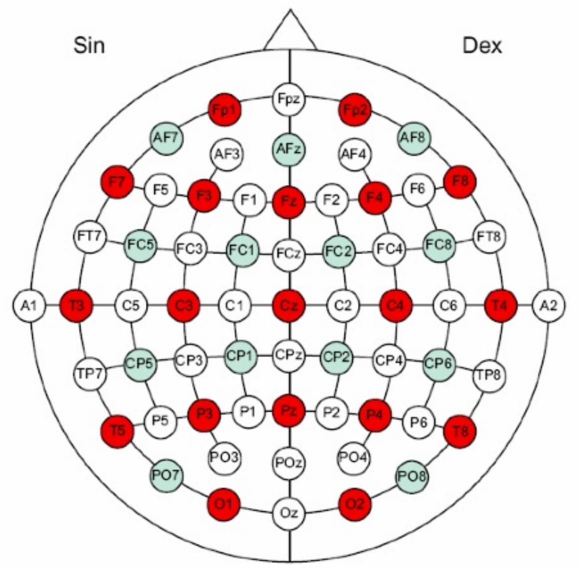


Figure 2.4: International system extended 10/20

There are three connection types: *unipolar*, *bipolar*, and *referred to the media.*
In *unipolar connection* one electrode is taken as a common reference for all others. The track takes the initials of the electrode from which it originated.
In *connection refers to the average* all electrodes are connected by a resistor network. The output of the network is taken as a reference. Also in this case the track assumes the initials of the electrode from which it originated.
Finally, in *bipolar connection* do not use any reference. Using the voltage differences between pairs of electrodes. This allows a more precise localiza-

tion of the areas of activity. In this case the track assumes the initials of the two electrodes on which is made the difference.

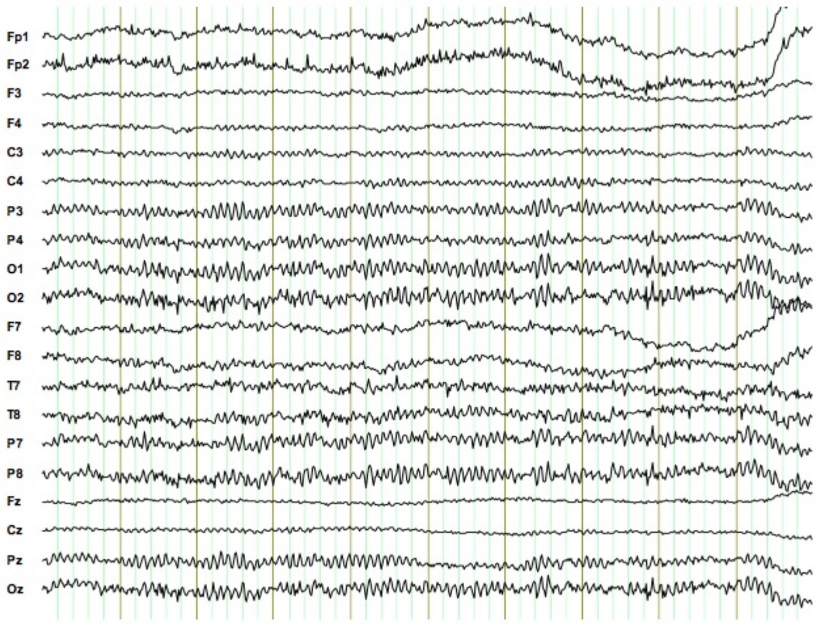


Figure 2.5: Unipolar EEG

2.2.3 Special EEG techniques

Special techniques complement the normal EEG recording scheme. Continuous EEG (cEEG) means recording EEG continuously for extended periods of time. This leads to challenges in electrode design and in data storage methods. The electrodes should be designed so that their impedance levels do not deteriorate too much during long recordings. Patient safety also becomes a concern as skin contact is maintained for several hours or even days. However, evidence suggests that cEEG is necessary to provide an accurate picture of the patients state.
In video EEG, a camera is used for recording video simultaneously with the EEG. Video can help clinicians identify artefacts. It provides behavioural correlates with EEG. Using depth electrodes to record electrocorticogram (ECoG) can provide a less distorted and artefact-free signal and thus facilitate interpretation. ECoG is, however, an invasive modality.

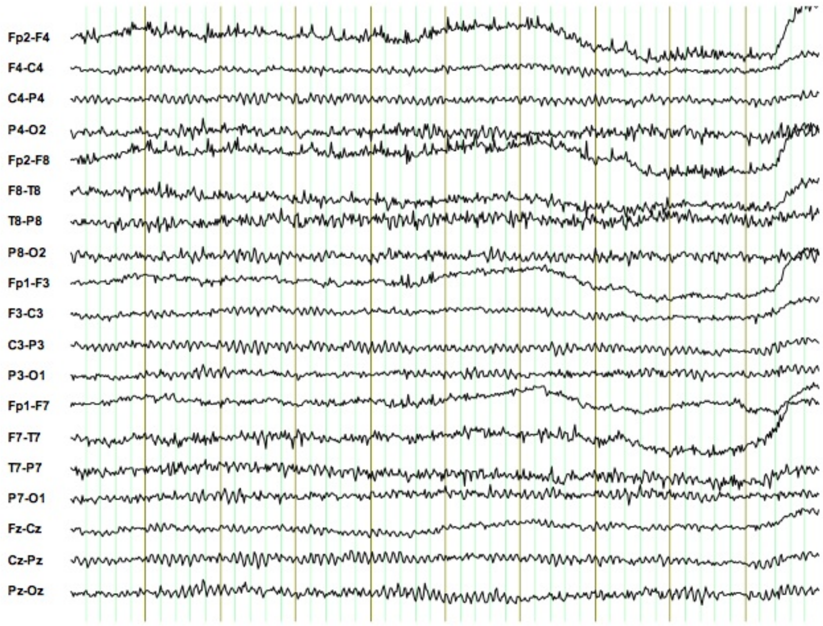


Figure 2.6: Bipolar EEG

2.3 Common artefacts

The goal of EEG monitoring is to infer the neurological state of the patient. Artefacts can contaminate the measurement and even lead to a false diagnosis. Artefacts can be divided into **physiological** and **mechanical** artefacts. Electromyogram (EMG), electrooculogram (EOG), electrocardiogram (ECG), ballistic effect, and glossokinetic potential are examples of physiological artefacts. EMG is caused by muscle activity and is typically manifested in high-frequency predominant activity. EOG reflects the movement of eyeballs. Since there is a voltage between the cornea and the retina, the eyeball acts like a dipole, also contributing to the scalp potential. EOG is best seen in frontal electrodes. ECG is caused by electrical activity of the heart. The ballistic effect overlays a pulse-synchronized signal on EEG due to pulsation of blood.

Typical mechanical artefacts are mains interference, mechanical movement of the electrodes and bed vibrations. Moving electrode leads also cause an artefact.

Extreme care must be taken to ensure that no wrong decisions are made because of compromised signal quality.

Chapter 3

Seizure Detection: State of the Art

The recording of seizures and spikes is of primary importance in the evaluation of epileptic behavior, but it is not easy because these events can be rare and unpredictable. Since 1970 researchers have developed methods for the automatic detection of spikes and, more recently, of seizures. The problems are complex because spikes and seizures have varied morphologies and they are not clearly defined.
The automatic detection methods can be of great assistance during long-term monitoring of epileptic EEG because they save a lot of time in the interpretation of long recordings. Every method developed requires human validation and no method is absolutely fail-safe. Recent developments include detection of the patterns specific to newborns and the possibility of warning a patient or observer that a seizure is starting.

3.1 Performance Measures:

Evaluation of the various methods presented in the following section is carried out by comparing the results of the corresponding algorithm (*Test outcome*) with the seizures marked by a neurologist (*Gold standard*).

		Condition (Gold standard)	
		True	False
Test outcome	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Quantitative analysis is performed using several performance measures. The most common used ones are:

- **Sensitivity (SE)**: ratio between the number of seizure detected and the total seizures, it is expressed in percentage.

- **Specificity (SP)**: ratio between the number of true negative decision and the total negative case, it is expressed in percentage.

- **False Positive Rate (FPR)**: refers to the number of times the detector declared the false seizure during one hour, measured in FP/h.

- **Latency**: delay between the expert marked seizure and the automatic detected seizure, expressed in seconds.

3.2 Main Articles

A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device [15]

Authors: H. Qu, and J. Gotman.
Year of publication: 1997
Method: a patient-specific classifier that uses five features extracted from the time and frequency domains. The features used for each epoch and for each EEG channel are: average wave amplitude, average wave duration, coefficient of variation of wave duration, dominant frequency, and average power in a main energy zone. The first three come from the waveform decomposition method of Gotman and the other two represent important characteristics of the EEG in the frequency domain. A modified nearest-neighbor (NN) classifier was used.
Data Set: 47 seizures in 12 patients.
Results:
Sensitivity : 100%, FP/h: 0.02/h, Latency: 9.35 s.
Conclusions: This method is able to detect seizure onsets accurately, but it has an important limitation because it is patient-specific and it only detects seizures similar to the template.

A system to detect the onset of epileptic seizures in scalp EEG[17]

Authors: M. E. Saab, and J. Gotman.
Year of publication: 2004

Method: Wavelet decomposition, feature extraction and data segmentation were employed to compute the a priori probabilities required for the Bayesian formulation used in training, testing and operation.
Data Set: 652 h of scalp EEG, including 126 seizures in 28 patients for training set and 360 h of scalp EEG, including 69 seizures in 16 patients for testing.
Results: (before /after tuning)
Sensitivity : 77.9%/ 76%, FP/h: 0.86/ 0.34, Latency: 9.8/ 10 s.
Conclusions: this system performed very well in every area and the sensitivity measured can be considered accurate because the algorithm is tested on data independent from the training one.

The tuning mechanism provided an important reduction in false detections with minimal loss of detection sensitivity and detection delay.

Seizure detection: evaluation of the Reveal algorithm[26]

Authors: Scott B Wilson, Mark L. Scheuer, Ronald G. Emerson, Andrew J. Gabor.
Year of publication: 2004

Method: Reveal algorithm utilizes three methods novel in their application to seizure detection: Matching Pursuit, small neural network-rules and new connected-object hierarchical clustering algorithm.
In this paper other two methods, *Sensa* and *CNET* are tested and compared with Reveal algorithm.
Data Set: 672 seizures from 426 epilepsy patients.
Results:
Reveal: Sensitivity :76%, FP/h: 0.11/h;
Sensa (Stellate): Sensitivity: 35.4%, FP/h: 0.11/h;
CNet: Sensitivity: 48.2%, FP/h: 0.75/h;
Conclusions: This study shows that Reveals performs better results then other methods.

Patient-specific seizure onset detection[20]

Authors: A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. Ted Treves, and J. Guttag.
Year of publication: 2004

Method: it uses a wavelet decomposition to construct a feature vector that captures the morphology and spatial distribution of an electroencephalographic epoch, and then it determines whether that vector is representative of a patients seizure or nonseizure electroencephalogram using the support vector machine classification algorithm.
Data Set: tested on noninvasive electroencephalograms from 36 pediatric subjects suffering from a variety of seizure types..
Results:
Sensitivity : 94%, FP/h: 0.25/h, Latency: 8.0 ± 3.2 s.
Conclusions: This method is patient-specific but slow (for 30 patients in 35 minutes), so the algorithm has not been tested on long term EEG moni-

toring.

Comparison of Fractal Dimension Estimation Algorithms for Epileptic Seizure Onset Detection[23]

Authors: G.E. Polchronaki, P. Ktonas, S. Gatzonis, A. Siatouni, H. Tsekou, D. Sakas. K.S. Nikita.
Year of publication: 2008

Method: it uses Fractal Dimension (FD) as natural measure of irregularity of a curve and it compares two methodologies: Katz's Algorithm and kith Nearest Neighbor Algorithm. If the FD value of a given window was below the threshold value, then that window was candidate of being the start of a seizure.
Data Set: 244.9 hours of scalp EEG recordings, including 16 seizures in 3 patients with refractory mesial temporal lobe epilepsy (MTLE).
Results:
Sensitivity : 100%, FP/h: 0.85/h, Latency: 6.5 s.
Conclusions: It is general as it does not require any prior information about the signal analyzed and it does not depend on pattern recognition as other works in the seizure detection literature do. Nevertheless, this results can only be compared to those of studies based on scalp EEG and the data set is very small and limited on patient suffering MTLE.
This method needs to be tested on more patients, more epilepsy types and with a separation of testing and training data set for threshold selection.

Automatic Detection of Seizure Onset in Pediatric EEG[10]

Authors: Y. U. Khan, O. Farooq, and P. Sharma.
Year of publication: 2012

Method: This paper proposes a method based on two statistical features: skewness and kurtosis with a wavelet based feature: normalized coefficient of variation (NCOV) were extracted from the data. The classification between seizure and normal EEGs was performed using simple linear classifier and the classifier declares the seizure if it was present in al least 60% channels.
Data Set: 55 seizures of 10646 seconds duration from 10 patients.
Results:
Sensitivity : 100%, FP/h: 1.1/h, Latency: 3.2 s.

Conclusions: This method performs onset detection of all the seizures with low latency but the dataset is very low.

3.3 Choise of Methods

In the previous chapter, we have seen that many different approaches have been experimented in order to produce EEG features that are specific to seizure and sensitive and generic enough to capture the majority of them.
Seizure detection has been approached as a machine-learning problem consisting in two main steps: generating features and designing a classifier.

The first step aims at compressing data and highlighting relevant events that describe signal.
When the features have been generated, it remains to define the decision making method. This is a typical problem of supervised learning. Given a development data set with known desired outcomes, one should design a system that performs well in the development data. Furthermore, the classifier should be able to generalize and show good performance also in previously unknown evaluation data set. Annotations made by a neurologist are often considered as the truth in the problem setting.
Machine-learning methods that have been applied in this field include expert systems, decision trees, clustering algorithms, self-organizing maps, and a variety of artificial neural network configurations.

This thesis proposed a comparison between two methods of feature extraction, more thoroughly presented in the following chapter, and investigate the approach of Artificial Neural Network (NN) for the decision making.
We have established certain goals during this project of thesis in collaboration with Micromed: the system developed should be reliable in terms of specificity and sensitivity and it must be able to detect both isolated seizures and prolonged ictal activity.

Chapter 4

Data acquisition and preprocessing

4.1 Data set

The data used for this project have been collected by Ing. Marco Cursi at St. Raffaele Hospital, Milano Italy.

The EEG signals included in the study belonged from different patients affected by epileptic seizure, a total of 30 EEG, with 58 seizures, datasets sampled at 256 Hz. The spectral bandwidth ranged from 0.5 Hz to 85 Hz.

In this project we have dataset with unipolar connection but it is simple to convert to bipolar.

4.2 Patients

The EEG signals belong to 30 patients aged between 13 and 93 years, including 13 women and 17 men, divided into different hospital wards:

Patient n.	Sex	Age	n. Epochs	n. Seizure
2	F	82	120	2
6	F	38	687	2
12	M	50	250	3
15	F	37	1400	6
18	M	47	234	1
20	M	44	312	1
22	F	32	5400	2
23	M	54	352	4
32	M	93	273	2
40	M	13	363	3
46	F	25	820	2
48	F	53	312	3
49	M	62	300	1
50	F	35	2105	2
51	F	67	500	3
52	M	74	1053	1
53	M	52	850	2
54	M	37	1200	1
56	F	48	698	1
60	M	54	1312	2
62	F	28	4370	3
66	M	85	358	1
70	F	78	688	2
71	M	43	910	1
79	M	65	802	1
81	M	51	415	1
82	F	54	390	1
83	F	53	1155	2
85	M	40	510	1
87	M	71	605	1

4.3 Acquisition System

For the acquisition of EEG signals, the Micromed device used at St. Raffaele Hospital is **SAM32 Acquisition Headbox** shown in Figure 4.1.
The SAM headboxes are a sophisticated and flexible acquisition system for EEG and polygraphy signals. They are the evolution of previous electroence-

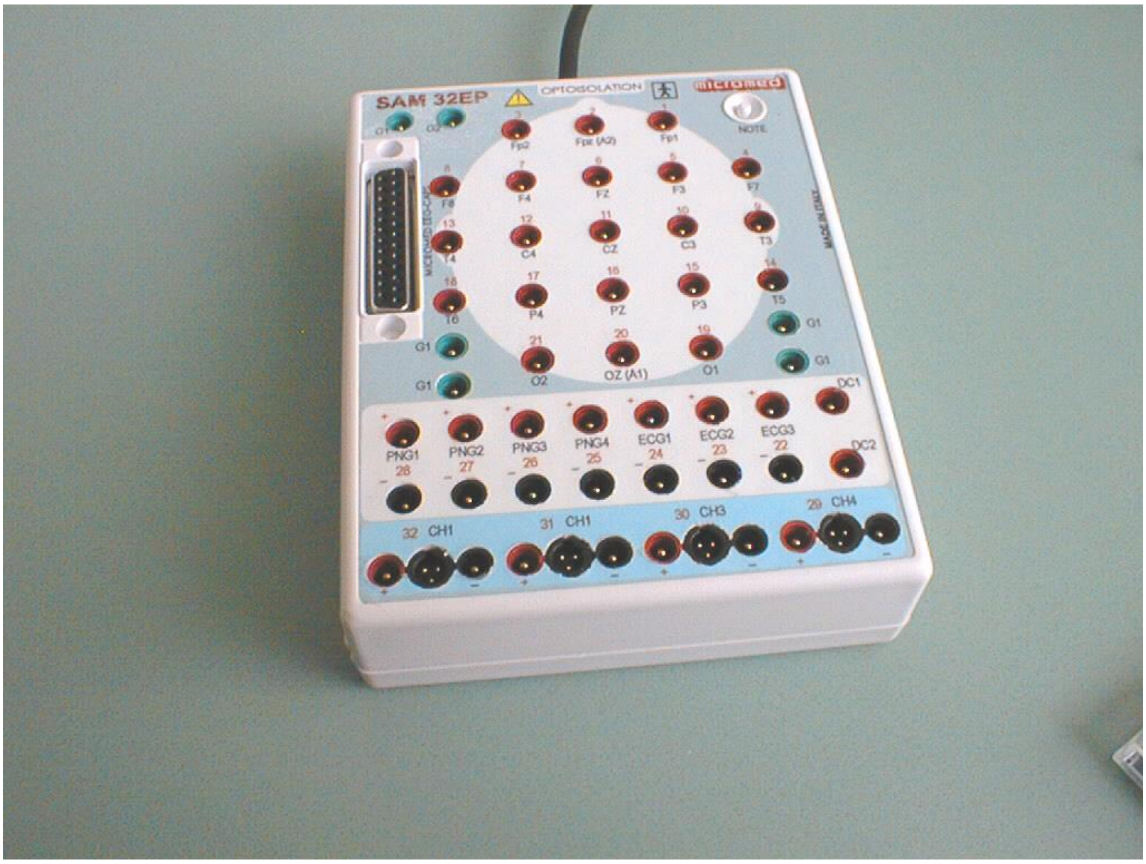


Figure 4.1: SAM 32EP Acquisition Headbox

falography and evoked potentials acquisition systems produced by Micromed that have been on the market for quite some time.

A SAM headbox is the part applied to the patient of an electroencephalography and evoked potentials system. It contains the amplifiers needed for the exam performance and the logical circuits for data transmission and reception. The device is physically isolated by a separation barrier composed by optoisolators and DC/DC converters from the PC to which it must be connected for data analysis.
It is SAM 32EP connected on one side to PC for the analysis of data and on the other side to the patient for the signal acquisition.

The interface to the PC is obtained through the software **SYSTEM PLUS**, that allows to manage the visualization, analysis and the storage of the traces, and the interface card **BQ USB** or **BQ USB**.

Figure 4.2: BQ USB

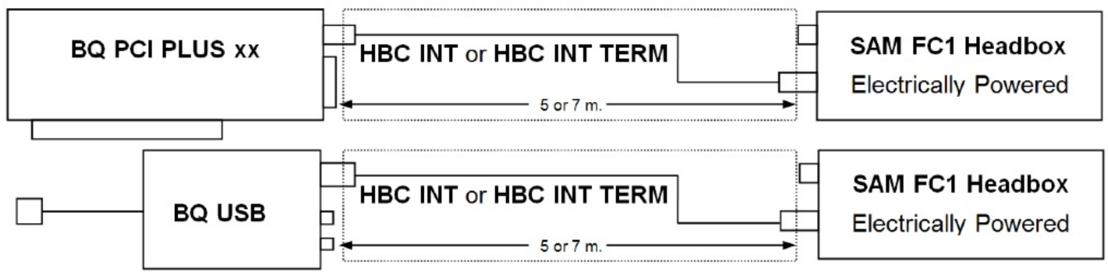


Figure 4.3: Micromed Devices

4.4 Filtering

Data were filtered with an analog Butterworth low-pass filter of order 3 with a cutoff frequency of 40 Hz.
The Figure 4.4 shows the frequency response of the filter, which, as we can see, does not eliminate but only attenuates the noise at 50 Hz. In order to completely eliminate noise contribution, the data were also filtered with an

IIR notch filter with cutoff frequency at 50 Hz.

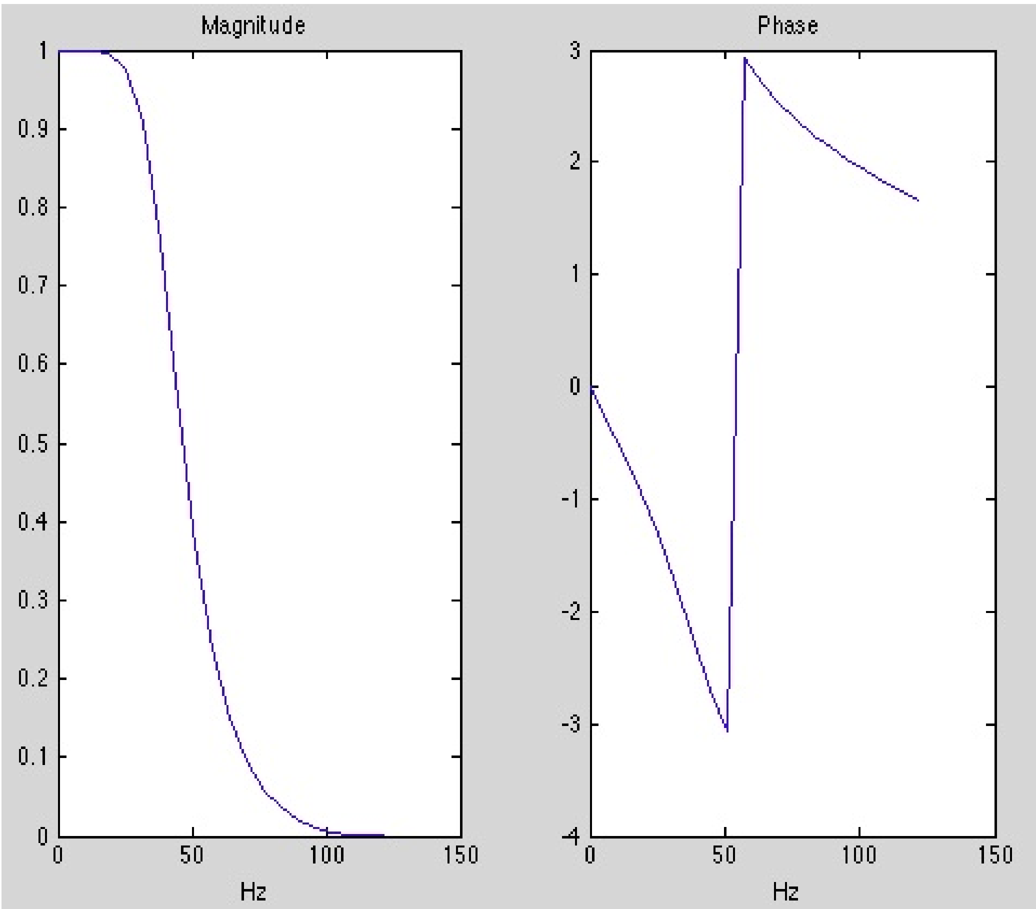


Figure 4.4: Butterworth Low Pass Filter

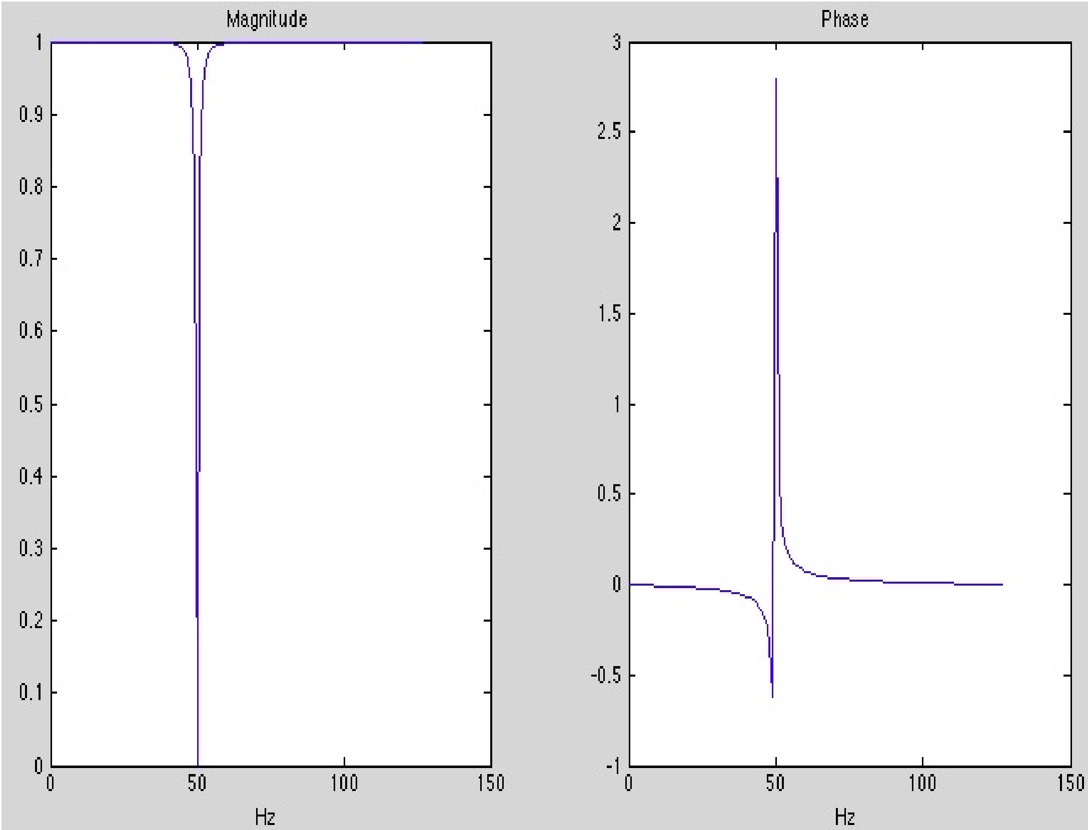


Figure 4.4: IIR Notch Filter

The frequency response of this filter is monotonic, and the sharpness of the transition from the passband to the stop-band is dictated by the filter order. Butterworth filter is the best compromise between attenuation and phase response, it rolls off more slowly around the cutoff frequency than the Chebyshev filter or the Elliptic filter, but it has no ripple in the pass band or the stop band.

4.5 Epoching

The EEG recordings are segmented into 2 seconds 50% overlapping epochs, which are long enough to capture the EEG characteristics to permit detection of 4 seconds seizures. A detection will be triggered when at least 3 epochs has a detected seizure status. Overlapping of the epochs is 50 %, therefore the minimum total seizure time needed for a detection is 4 seconds, that is exactly the length of the shortest seizure in the database.

We have chosen to remove seizures shorter than 4 seconds and to consider just one seizure if interictal event between two seizures lasts less than 2 seconds.

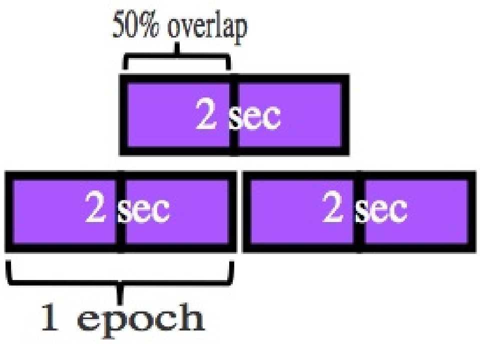


Figure 4.5: EEG recordings are segmented into 2 seconds 50% overlapping epochs

Chapter 5

Feature Extraction Methods

The detection system consists of three basic steps:

- **Selection**: choice of the important features to describe the model, the selection of parts of a system that must be taken in consideration depends essentially by the objective (goal);
- **Separation**: the boundaries that define the two classes to be identified must be defined with precision;
- **Parsimony**: among all the possible models that represent the system, the simplest is to be considered, according to the principle of Occam (Occam's razor): 'The simpler of two models, When Both are consistent with the Observed date, is to be preferred'.

Here, two different methods were applied to extract the information from EEG signals: Wavelet and AR coefficients, the Artificial Neural Network were then used as mean to classify seizures or not seizures pattern.
In this chapter there is a brief introduction to these methods implemented in the algorithm.
In the next chapter discusses in detail the behavior of neural networks and the choices made in the design of the network. I decided to dedicate an entire chapter because most of my project was based on the study of the function and methods to optimize the classification of seizure by using this method.

5.1 Wavelet Transform (WT)

WT provides a better compromise between time-resolution and frequency-resolution compared to other techniques of time-frequency analysis, such as those based on Short Time Fourier Transform (STFT). The latter, in fact, has the disadvantage of maintaining a given fixed time-window for all frequencies of the analyzed signal: this may causes a substantial loss of information at very low frequencies or very high frequencies. The basis of this phenomenon is the Heisenberg's Uncertainty Principle, which determines that the increase in time resolution leads to a loss in frequency resolution and vice versa. Wavelet analysis of the EEG signal allows to avoid this problem by using of longer time intervals for higher frequency resolution at low frequencies and shorter time intervals for more precise higher time resolution at high frequencies.

In Figure 5.1 is schematized the time-frequency resolution for the STFT and WT:

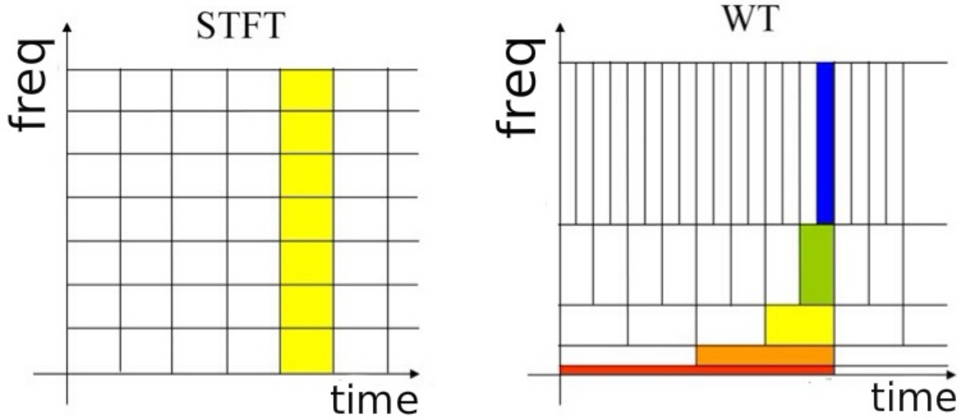


Figure 5.1: Time and frequency resolution of STFT (Δt=cost and Δf=cost)and WT (Δf/f=cost)

The Wavelet Transform decomposes the signal of interest by a set of basis functions, called Wavelets.
Each of these basic functions is a scaled and translated version of a function prototype, the **Mother Wavelet**. The basic functions in the frequency domain are the frequency responses of band-pass filters, whose width of the passband is determined by the temporal length of the window.
Therefore, WT uses windows of variable length: short at high frequencies and long at low frequencies. The main advantage of wavelet transform is that it allows multiresolution analysis to optimize the temporal resolution and the frequency for each frequency value.

Interpreting the WT as a filter bank, the width of the bandwidth varies with its center frequency according to the law:

$$\frac{\Delta f}{f} = Q = cost \tag{5.1}$$

The filter bank will therefore composed by band-pass filters with constant relative bandwidth (Q-factor constant). When the equation 5.1 is satisfied, then Δf and Δt change with the central frequency of the filter in question. Therefore the frequency resolution (and hence in time) depends on the frequency in question. For the Heisenberg Uncertainty Principle, the time and frequency resolutions can not be both arbitrarily small, since their product is bounded below by:

$$\Delta f \Delta t \geq \frac{1}{4\pi} \tag{5.2}$$

Even in the case of WT, obviously, that principle must be satisfied but due to the multiresolution property that characterizes them, the time resolution becomes arbitrarily good at high frequencies and the frequency resolution becomes arbitrarily good at low frequencies.

In 2005 wavelet transform method and short time Fourier transform method were compared to determine their accuracy to determine the epileptic seizures, and wavelet transform resulted to give superior performance [12].

Continuous Wavelet Transform

Continuous Wavelet Transform (CWT) follows exactly the criteria mentioned above but introduces a simplification: all the impulse responses of the filter bank are defined as a translated and scaled version of the same prototype or mother wavelet, h (t):

$$h_{a,\tau}(t) = \frac{1}{\sqrt{|a|}} h\left(\frac{t-\tau}{a}\right) \tag{5.3}$$

where a is a scale factor, the constant $1/\sqrt{|a|}$ is used for the energy normalization and τ defines the translation.

The Continuous Wavelet Transform of a signal $x(\tau)$ is defined as:

$$CWT_x(\tau, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) h^* \left(\frac{t-\tau}{a}\right) dt \tag{5.4}$$

where h^* is the complex conjugate of the wavelet mother function and $h^*\left(\frac{t-\tau}{a}\right)$ is a version traslated and scaled. For each instant of time t, the CWT calculates the correlation between the signal x(t) and a version translated and scaled of h^*.

Select an appropriate type of wavelets for a given application involves finding a trend of mother wavelet that matches the features of interest of the analyzed signal. In the case of the analysis EEG the choice of a mother wavelet with characteristic oscillatory activity is appropriate [Allen et al., 2009]. The complex Morlet Wavelet, defined by the following expression, can be considered a good choice:

$$h(t) = \frac{1}{\pi^{1/4}} exp(j2\pi f_0 t) exp\left(-\frac{t^2}{2}\right) \tag{5.5}$$

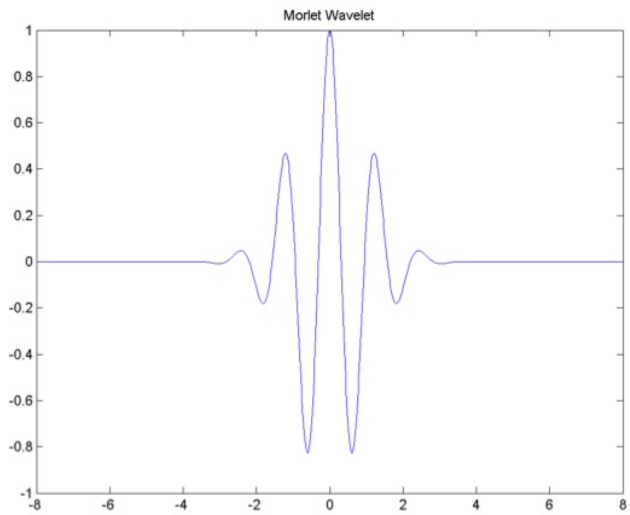


Figure 5.2: Morlet Function

The Morlet CWT function is one of the most popular types of wavelet functions in the biomedical field due to its simplicity and suitability for spectral estimations [25].

Algorithm

From 5.4:

$$CWT_x = \frac{1}{\sqrt{|a|}} \sum_k \int_k^{k+1} x(t) h^* \left(\frac{t-\tau}{a} \right) dt$$

since $x(t) = x(k)$,if $t \in [k, k+1]$ then

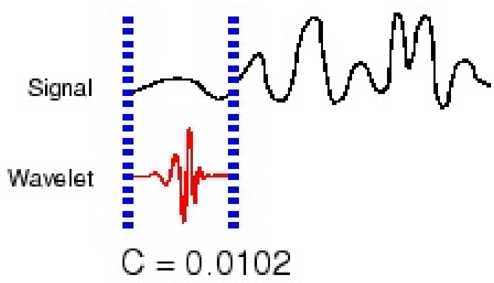
$$CWT_x = \frac{1}{\sqrt{|a|}} \sum_k x(k) \left(\int_{-\infty}^{k+1} h^* \left(\frac{t-\tau}{a} \right) dt - \int_{-\infty}^{k} h^* \left(\frac{t-\tau}{a} \right) dt \right)$$

so at any scale a, the wavelet coefficients CWT_x for $\tau = 1$ to *length(x)* can be obtained by convolving the signal x and a dilated and translated version of the integrals of the form $\int_{-\infty}^{k} h^*(t) dt$ and taking finite difference [30].

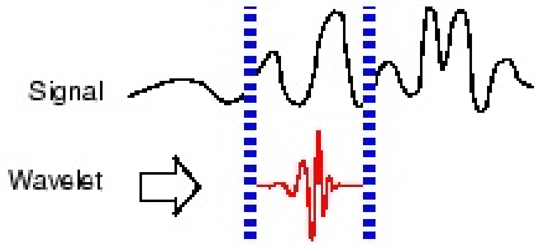
The basic algorithm for CWT could be described by the following steps:

1. Consider a wavelet and compare it with the starting section of the original signal.

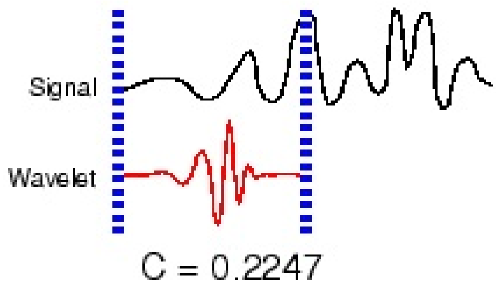
2. Calculate a number, $C = CWT_x$, that represents how closely correlated the wavelet is with this section of the signal. The larger the number C is, in absolute value, the higher is the similarity. This follows from the fact the CWT coefficients are calculated with an inner product, that give a measure of similarity.
   As described in Definition of the Continuous Wavelet Transform, the CWT coefficients explicitly depend on the mother wavelet. Therefore, the CWT coefficients are different when computed on the same signal with different wavelets.



3. Shift the wavelet to the right and repeat steps 1 and 2 until the whole signal is covered.



4. Scale (stretch) the wavelet and repeat steps 1 through 3.



5. Repeat steps 1 through 4 for all scales.

Multiplying each coefficient by the appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal.

5.2 AR Modeling

In addition to Frequency-based and time-based methods previously developed for seizure detection, for example Fourier and Wavelet transformation, the Auto Regression (AR) model is widely used to convey the spectral information.
The AR parameters suppress the noise effect and emphasize the characteristics of the signal while FFT process the signal and noise equally.

For a process of identification of a suitable model that can describe the data of biological interest the basic steps are:

- observation data
- choice of model and estimate optimal order
- estimation of the model parameters
- Model Validation

Autoregressive model for EEG

Assuming that EEG signal $x(n)$ has a zero mean and may be considered as a stationary signal in a finite time window of 2 seconds, it is possible to represent the current observation $x(n)$ as a linear combination of past values and white noise as formulated in Eq. 5.6:

$$x(n) = \sum_{k=1}^{M} a_k x(n-k) + e(n) \tag{5.6}$$

where a_i are AR parameters and $e(n)$ is white noise with zero mean.
In the current study the AR parameters are estimate based on the Yule-Walker equation utilizing the Least Mean Squared (LMS) method criterion. LMS determines an estimate of the parameters defining an index J equal to the mean squared error and then finding the minimum of J based on the parameters.

The optimal order for the AR model is not known a priori, but must be evaluated from the data. Increasing the order of the model, and therefore its complexity, the model best approximates the data. The order p can be estimate by three criteria:

- **Mean Squared Error (J)**: stops the order p when the improvement introduced by the model of order p +1 is not significant, when it is less than a predetermined value ϵ

$$\frac{J_{min}^{p+1}}{J_{min}^{p}} \geq 1 - \epsilon \tag{5.7}$$

- **Akaike Information Criterion (AIC)**: minimizing the index

$$AIC = ln(J_{min}^{(p)}) + \frac{2(p+1)}{N}, \tag{5.8}$$

where N represent the number of samples in $x(n)$.

- **Final Prediction Error (FPE)**: minimizing the index

$$FPE = \frac{N+p-1}{N-p-1} J_{min}^{(p)}, \tag{5.9}$$

This criterion represents a trade-off between the estimation error and the size of the model order. AIC and FPE are inclined to overestimate the optimal order, especially for large N.

The results obtained with these methods are shown in the chapter of Implementation, using a representative example.

Chapter 6

Classification Method: Neural Network

Neural networks offer a very powerful set of tools that allows solving problems in the context of classification, regression and control of non-linear. The inspiration for neural networks is derived from studies on mechanisms of biological information processing in the nervous system, particularly the human brain.

An artificial neural network (NN = Neural Network) is a processing system of informations that has features in common with biological neural networks. They have been developed as a generalization of mathematical models based on the following assumptions :

- Processing takes place in correspondence of simple elements called neurons;
- Signals pass between neurons through communication link;
- Each communication link is associated with a weight that, typically, multiplies the transmitted signal;
- Each neuron determines the output signal by applying an activation function, usually not linear, to the weighted sum of its inputs.

We will focus on a particular class of neural network models: the open-chain networks (*feedforward*). These networks can be viewed as non-linear mathematical functions that transform a set of independent variables $x = (x_1, ..., x_d)$, called **inputs** of the network, into a set of dependent variables $y = (y_1, ..., y_c)$, called **outputs** of the network.

The precise form of these functions is relative to the internal structure of the network and to a set of values $w = (w_1, ..., w_d)$, called **weights**. So we can write the function of the network in the form $y = y(x; w)$ which denotes the fact that y is a function of x parameterized by w.

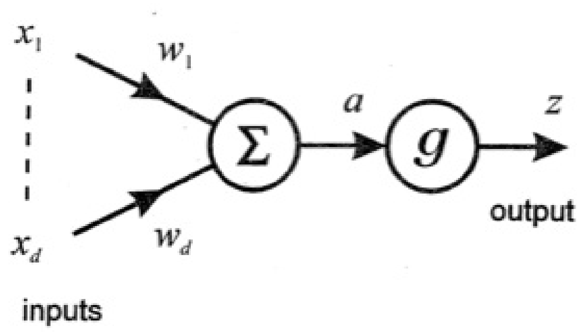


Figure 6.1: Artificial Neural Network

A simple mathematical model of a single neuron is represented in Figure 6.1 and it may be seen as a non-linear function that transform the input variables $x_1, ..., x_d$ in the output variable z.
In this model, the weighted sum of the inputs is performed, using as weights the values $w_1, ..., w_d$ (which are analogous to the powers of the synapses in the biological network), thus obtaining:

$$a = \sum_{i=0}^{d} w_i x_i \tag{6.1}$$

The values of weights may be of any sign, which depends on the type of synapses.

The output z is obtained by applying to a nonlinear transformation $g()$, called **activation function**, obtaining:

$$z = g(a) = g(\sum_{i=0}^{d} w_i x_i). \tag{6.2}$$

Some possible forms for the function $g()$ are represented in Figure 6.2.

Summing up, the neural network is characterized by the way in which its neurons are interconnected, that is its *architecture*, by the method used to

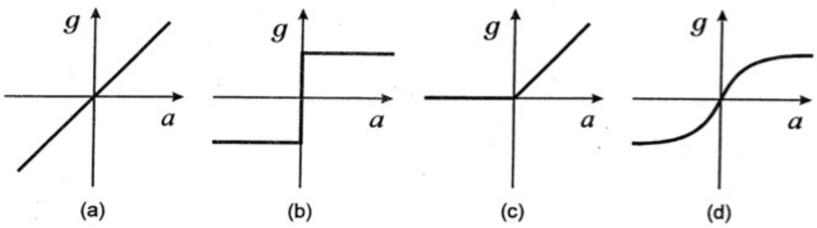


Figure 6.2: Some possible forms of activation function

determine the weights of the connections, *algorithm of learning* or *training* and by the *activation functions* of the neurons themselves.

A neural network is composed of simple processing elements called *neurons, units, cells* or *nodes.* Each neuron is connected to other neurons by means of communication link to which is associated a weight. The weights represent the primary information used in solving a particular problem and their value must be set properly during the training phase.

Each neuron has an internal state, called *activation* or *activation level*, a function of the inputs of the neuron itself; typically a neuron uses this function to construct the signal to be sent to the neurons to which there is a communication link output.

6.1 Activation Functions

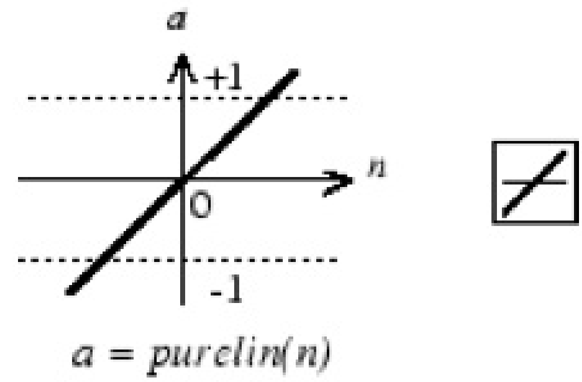
The primary operation of a single artificial neuron consists of a weighted sum of input signals and in the application of an activation function in output. Typically the same activation function is used for all the neurons belonging to a particular layer of the network.

The activation functions that are implemented in Matlab are:

- *Linear Transfer Function*:

$$f(x) = x, \forall x \tag{6.3}$$

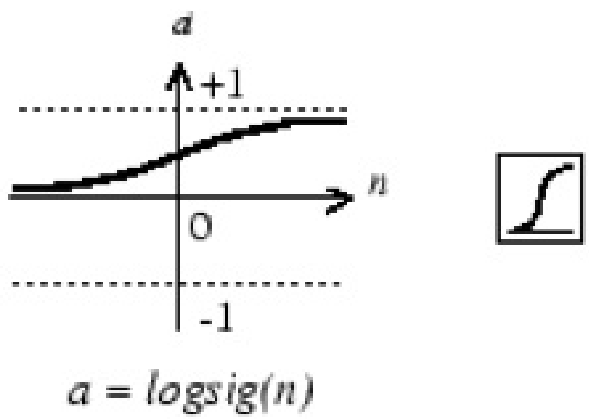
  the output of a neuron having this level of activation is simply equal to the weighted sum of the input signals;



- *Log-Sigmoid Transfer Function* with steepness factor σ:

$$f(x) = \frac{1}{1 + exp(-\sigma x)} \tag{6.4}$$

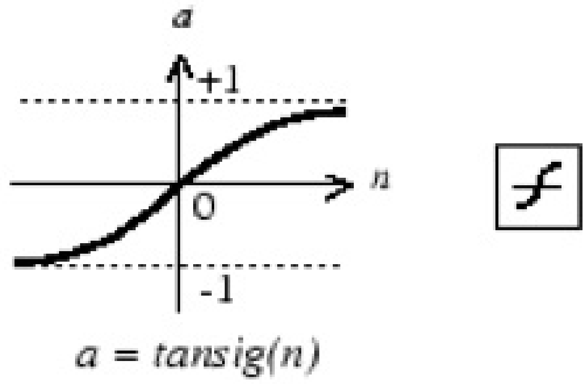
The function *logsig* generates outputs between 0 and 1 as the neurons net input goes from negative to positive infinity.



- *Tan-Sigmoid Transfer Function* with steepness factor σ:

$$f(x) = \frac{1 - exp(-\sigma x)}{1 + exp(-\sigma x)}, \tag{6.5}$$

this function gives an output between -1 and 1 and a steepness factor of $\sigma = 2$ corresponds to the hyperbolic tangent function, which is also widely used as a level of activation of a neuron.



The transfer function selected for the analysis on detection of seizures studied in this project is the Tangent Sigmoid because Sigmoid output neurons are often used for pattern recognition problems, while linear output neurons are used for function fitting problems.

6.2 Weight Adjustments

The way the values of weights are set (training) is another distinguishing factor of a particular neural network with respect to another.
There are two methods of training: *supervised learning* or *unsupervised.*
In this project we use supervised learning, in which the number of sequences (features) of input vectors is available, each of which corresponds to an output vector that represents the desired behavior of the neural network in the face of a particular input; once provided this information the network weights are adjusted using a particular training algorithm.

Before the use of the network, we need to identify the model, i.e. we have to determine all the parameters w. The process of determination of these parameters, as we have said, is called **training** and it can be an action very intense from the computational point of view. However, once we have defined weights, new inputs can be processed very quickly.

To train a network, we need a set of examples, called **training set**, whose elements are pairs (x_q, t_q), $q = 1, ..., n$, where t_q is the desired output value, called **target**, in correspondence of the input x_q.
The training consists in finding the values for the parameters w that minimize an appropriate **error function**. There are several forms of this function,

however, the most used is the *Sum of Squared Residuals*.
The residuals are defined as

$$r_{qk} = y_k(x^q; w) - t_k^q \tag{6.6}$$

The error function is

$$E = \frac{1}{2}\sum_{q=1}^{n}\sum_{k=1}^{c} r_{qk}^2 \tag{6.7}$$

where E depends on x^q and t^q, which are known values, and on w unknown. To find the minimum of the error function the easiest method is an iterative method based on the technique of **gradient descent**.

The plot 6.2 shows three lines because the data of the input and output vectors are randomly divided into three different sets:

- training set
- validation or testing set
- verification set

The 60 % of these data is used to train the network, the 20 % is used to verify that the network has learned to generalize. The training procedure in fact continuous at each iteration by reducing the error of the network on the validation vectors. When you reach the minimum variance mean square error, the network stores the weight matrix and stops the operation of training.

Thus the validation data set is used to decide when to stop the training, so as to avoid the so called "overtraining", which could make the network unable to identify new data belonging to the same classes of the training examples. The remaining 20 % of the data carriers previously considered is used to test the network, that is, to analyze the generalization ability of the network with the same input data that has never seen and then analyzed.

Each time a feed-forward neural network is initialized, the initial parameters are different and may produce different results.

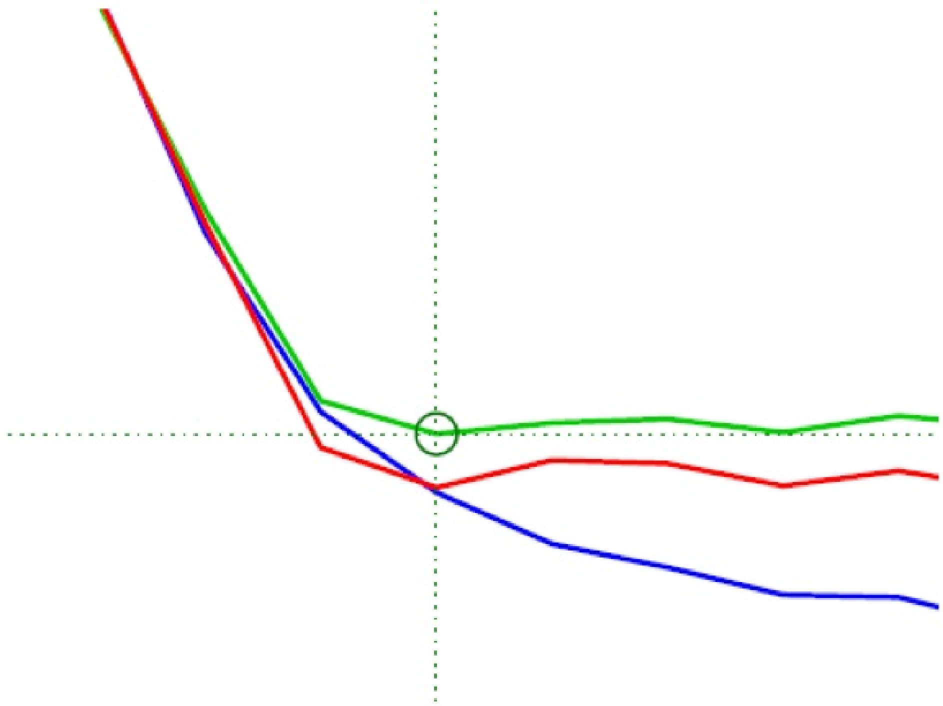


Figure 6.3: Performance Plot

6.3 Preprocessing and Postprocessing

The training of a neural network can be made very efficient if we perform a priori procedures for processing the inputs and outputs of the network.
The preprocessing routine that have been used are guaranteed automatically whenever a neural network is created.
The functions of preprocessing transform inputs into better shape for neural networks. The same is obviously applied to the output values of the network (targets) to optimize the training procedure.

The feedforward neural network used the features *fixunknowns*, *removeconstantrows* and *mapminmax*. For the output values instead *removeconstantrows* and *mapminmax*.

mapminmax transforms input data so that all values fall into the interval [-1, 1].

removeconstantrows removes the rows of the input vector that correspond to input elements that always have the same value, because these input elements are not providing any useful information to the network.

fixunknowns recodes unknown data (represented in the users data with NaN values) into a numerical form for the network. *fixunknowns* preserves information about which values are known and which are unknown.

6.3.1 Principal Component Analysis

In some situations, the size of the input vector is very large, but the components of the vectors are strongly correlated with each other (redundant).
For this reason it is useful in these cases, reduce the size of input vectors.
A procedure that allows to perform this task is the **principal component analysis**. This technique produces essentially three effects: it orthogonalize the components of the input vectors (so that they are uncorrelated among themselves), it order the orthogonal components thus obtained so that those with the largest standard deviation are positioned first, and those which contribute minimally the variation in the data set, are completely eliminated.

Principal component analysis is a technique not reversible; for this reason it is recommended only for the processing of the input. The outputs require processing techniques reversive.

As mentioned earlier, the behavior of a neuron is determined by its activation function and the pattern of weighted connections along which sends and receives signals; such information describes the architecture of a neural network.
The input units are not considered as a layer, as it does not perform any computation; equivalently the number of layers in a neural network can be defined as the number of layers of the communication link between the neurons.

In the Figures in the Chapter 8, there are the representations of the networks used in this project.
The former is used with in input the wavelet coefficients, the latter with in input 10 coefficients of AR model and the amplitude of the signal for each epoch.

Both have one hidden layers with an appropriate number of neurons and one output that could be 1 if the seizure is present or 0 otherwise, as in Figure 6.4.

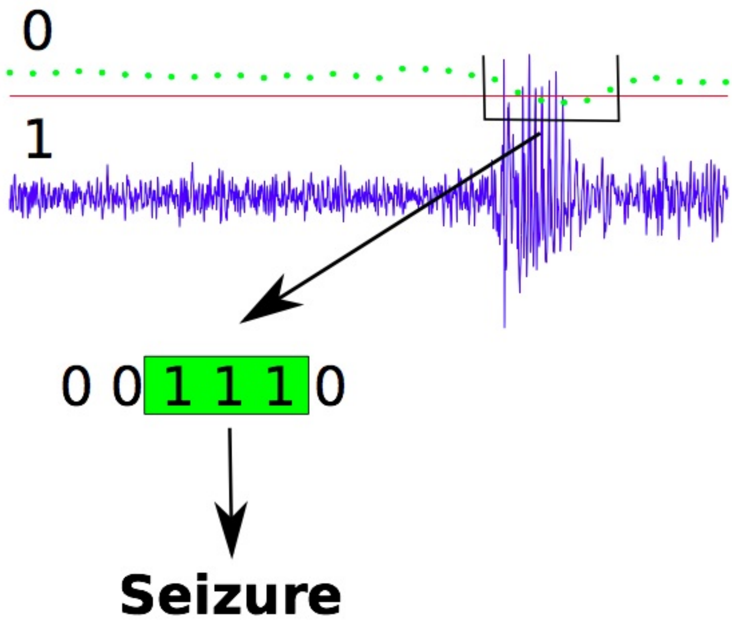


Figure 6.4: Binary Classification

The network uses the training algorithm *trainrp* function (Resilient Backpropagation), because it is the fastest algorithm on pattern recognition problems.

Chapter 7

Validation

During the project of a classifier, before using it, it is important to validate it, that is, to measure its performance through an estimate of the classification error. This is also very useful for comparing the performance of various classifiers.

The assessment of the performance of a neural network require a quantitative evaluation of the obtained solution: once completed the training, one must be able to provide a measure of the performance of the neural network.

The method used in the present work is the **Leave-One-Out Cross Validation** that uses a single database instance as a test set and the remaining as the training set. Training and tests are repeated until each element of the whole database is used once as a test. However, this technique is computationally expensive.

7.1 Leave One Out

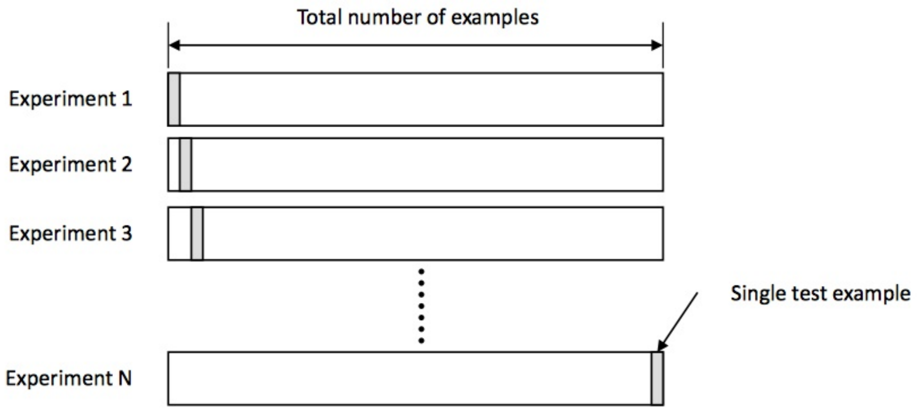
We evaluate the performance as the average of the performance achieved on each of the N fold:

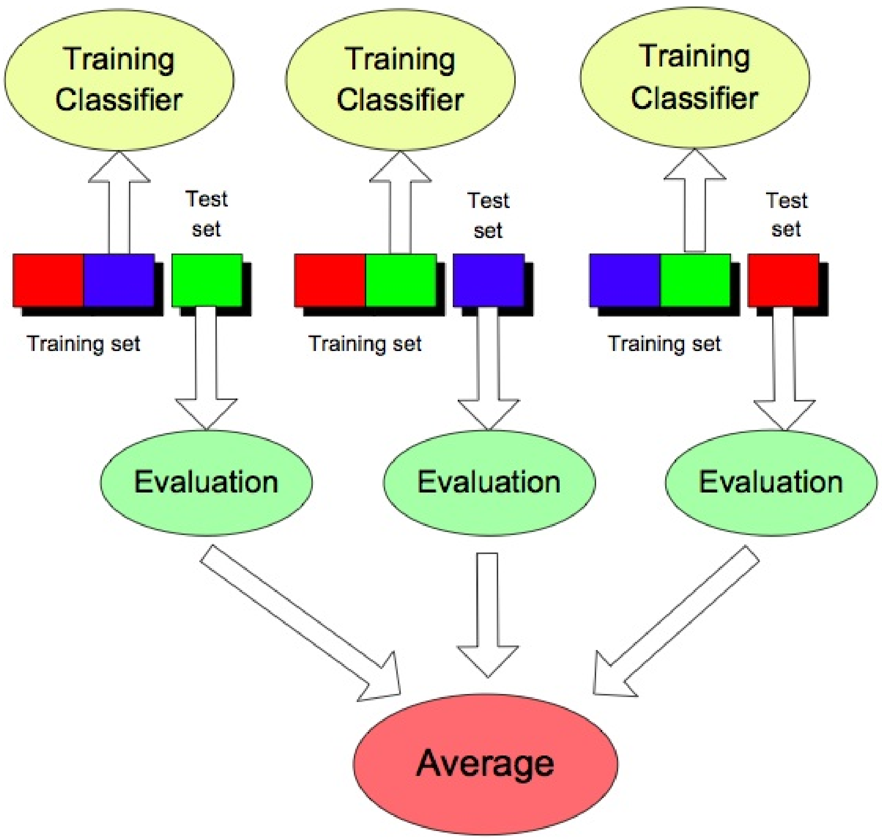
- For a dataset with N signals, perform N tests
- For each experiment use $N-1$ signals for training and the remaining signal for testing

- Calculate P_{err} as the ratio between the number of epochs classified in a wrong way, and the number of total epochs

$$P_{err} = \frac{\#error}{\#epochs}$$

- Evaluate the performance as the average of the performance achieved on each of the N fold





7.2 Binary classification Test

As it has already been mentioned in Chapter Three, statistical measures of the performance of a binary classification test are **sensitivity** and **specificity**.

Sensitivity, also called the *true positive rate*, measures the proportion of actual positives which are correctly identified as such.

Specificity (sometimes called *true negative rate*) measures the proportion of negatives which are correctly identified as such.

<table>
<tr><td colspan="2"></td><td colspan="2">Condition
(as determined by "Gold standard")</td></tr>
<tr><td colspan="2"></td><td>Condition positive</td><td>Condition negative</td></tr>
<tr><td rowspan="2">Test outcome</td><td>Test outcome positive</td><td>True positive</td><td>False positive
(Type I error)</td></tr>
<tr><td>Test outcome negative</td><td>False negative
(Type II error)</td><td>True negative</td></tr>
<tr><td colspan="2"></td><td>Sensitivity = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$</td><td>Specificity = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$</td></tr>
</table>

Figure 7.1: Binary Classification Test

In the next chapter we will present the results obtained by varying the number of hidden layers and nodes in the neural network, in particular the algorithm calculate the percentage of true positives (percentage of the number of epochs about seizure detected on the total number of epochs really a seizure) and true negatives (number of epochs of not seizure detected correctly).

7.3 Number of layers and neurons

In relation to the number of layers, there are numerous methodological contributions according to which a single hidden layer is sufficient to approximate the nonlinear functions with a high degree of accuracy:

UNIVERSAL APPROXIMATION THEOREM
A multilayer neural network with a single hidden layer is capable of learn any continuous function and approximate it with any degree precision.
(Kolmogorov existence theorem applied to neural networks by Cybenco,1988 and by Hornik,1989)

The Universal Approximation Theorem, however, has left an open problem, which is that of the choice of parameters: there is no method that allows us to know, a priori, what is the number of nodes to be placed in the hidden layer, or how many layers insert if the number of nodes is elevate, or what is the number of iterations needed in the training.

From the choice of these parameters depends the approximation and generalization ability and efficiency of the network: when there are fewer nodes in the middle layer, the higher the generalization ability of the network, but at the same time decreases the accuracy in the approximation.

Increasing the number of layers, increases the complexity of the situation that the network can understand, but grows the time required for learning, and the computational cost.
The greater is the number of iterations, the greater is the degree of precision that the network reaches, but the greater is the risk of overfitting, that is the network could remain too tied to the data and is not able then to give good feedback to situations never seen before.

For the Universal Approximation Theorem there are numerous empirical studies that estimate the predictive network with one hidden layer; the limit of this approach that it often requires a large number of neurons, which limits the learning process. The use of networks with two hidden layers is more effective for problems of prediction of high frequency data.

With reference to the number of neurons that must be assigned to each hidden layer, which rely on the criterion that minimizing the risk of overlearning that takes place when you decide to place an excessive number of neurons, which draw nearly perfectly the pattern of the time series but are not able to generate reliable forecasts because they reduce the contribution of the inputs.
Conversely, the risk of assigning a too low number of neurons is to reduce the potential for learning of the neural network. It is therefore necessary to find a compromise between too many or too few neurons.

The formulas proposed in the literature [3] are very different and, in some cases, contradictory:

$$n_{hl} = 2n_{input} + 1$$

$$n_{hl} = 2n_{input}$$

$$n_{hl} = n_{input}$$

$$n_{hl} = \frac{n_{input} + n_{output}}{2} + \sqrt{n_{training}}$$

where:
n_{hl} is the number of neuron in the hidden layer;
n_{input} is the number of input;
n_{output} is the output number;
$n_{training}$ is the number of observations in the training set.

The empirical results show that none of these rules is generalizable to any forecasting problem.

Chapter 8

Implementation

This chapter describes the algorithms that were developed in the course of this thesis.

8.1 Code Implementation

For the implementation of the codes MatLab program was used (**Mat**rix **Lab**oratory). It is an environment for the numerical calculation and statistical analysis which includes the eponymous programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran.
At the beginning it was decided to use an interactive Matlab toolbox, EEGLAB, in order to read the tracks of Micromed data and load them into the working environment MatLab; but for paths of size too big, this toolbox has many problems of memory management. We have decided to use the procedures developed directly in Micromed by Ing. Raffaele Orsato for the conversion of the tracks recorded with their equipment into signals readable by MatLab program and vice versa.

After the acquisition of EEG signal from the patients and the collection of the dataset, described in *Chapter 4*, the principal steps of the algorithm implemented are shown in the Figure 8.1

Now, a representative study will be presented to explain the algorithms developed.

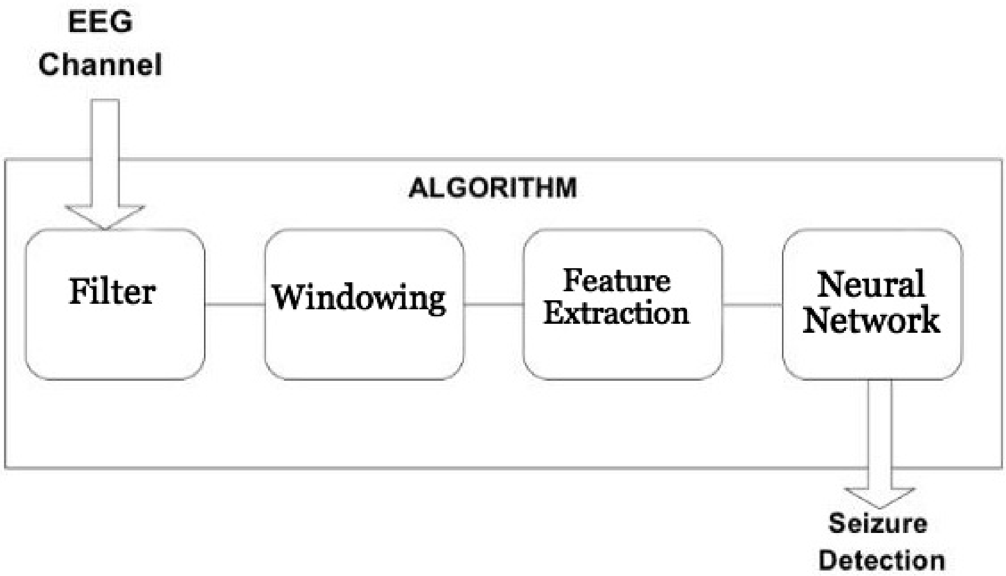


Figure 8.1: Schematic representation of the operation of the algorithm

8.2 Preprocessing

For the step related to the preprocessing, as has already been explained in *Chapter 4*, filtering was performed using a Butterworth low-pass filter and a notch filter to permanently delete the noise component at 50 Hz.

An example of data belonging to the database is shown in the following figures, with a representation of power spectrum before and after the filtering:

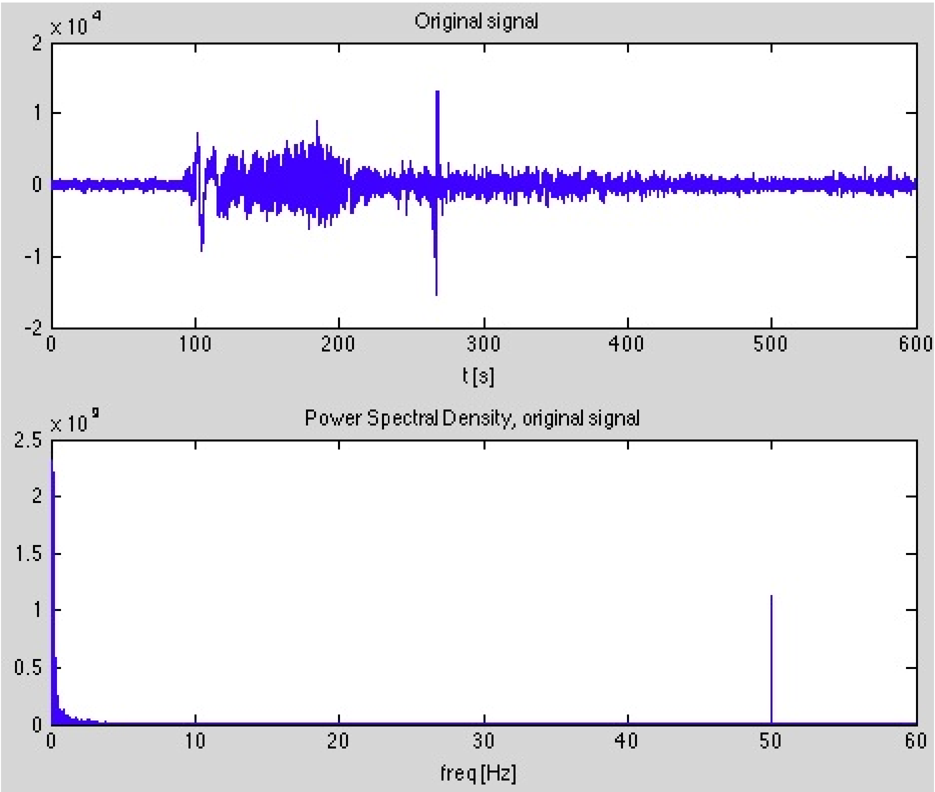


Figure 8.2: An example of EEG signal, affected by noise, and its power spectrum

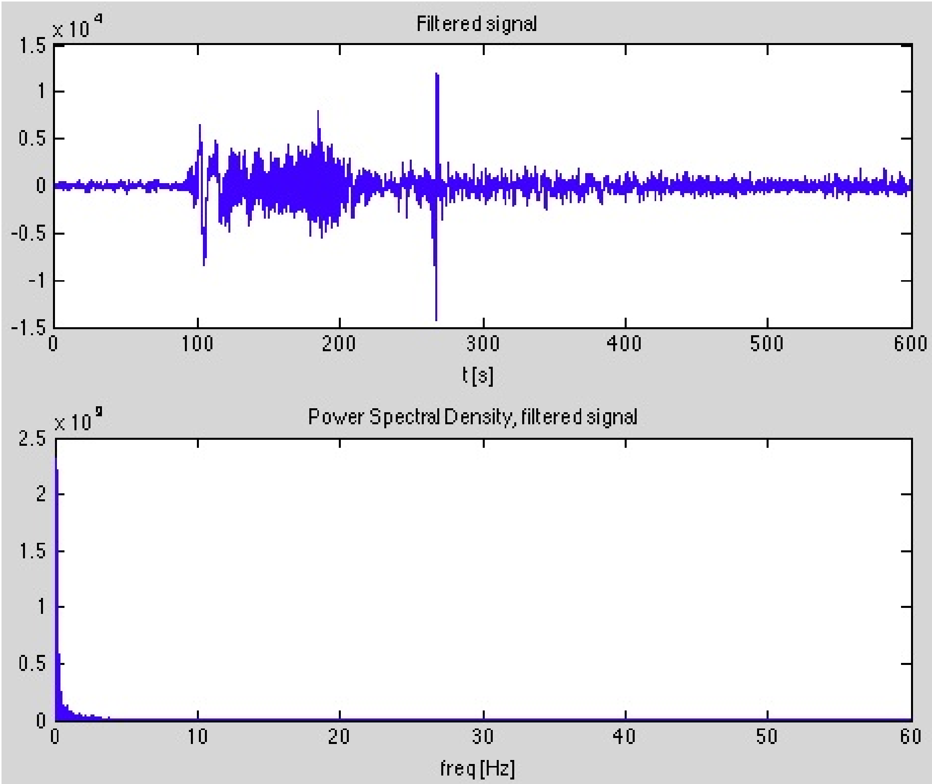


Figure 8.3: An example of filtered EEG signal and its power spectrum

8.3 Wavelet Transform

For each EEG signal in the database is performed the wavelet transformation in the individual epochs by using the Matlab:

$$C = cwt(x, scales,' morl');$$

The operation of this function is explained in detail in *Chapter 5* and the scales used are between 11 and 64, that correspond to frequency between 3 and 19 Hz (see Matlab function $scal2frq$).

During epileptic events we can easily notice an increase in power shown visually by the wavelet coefficients, which give us information about the frequency you have chosen to test (scale) and time.

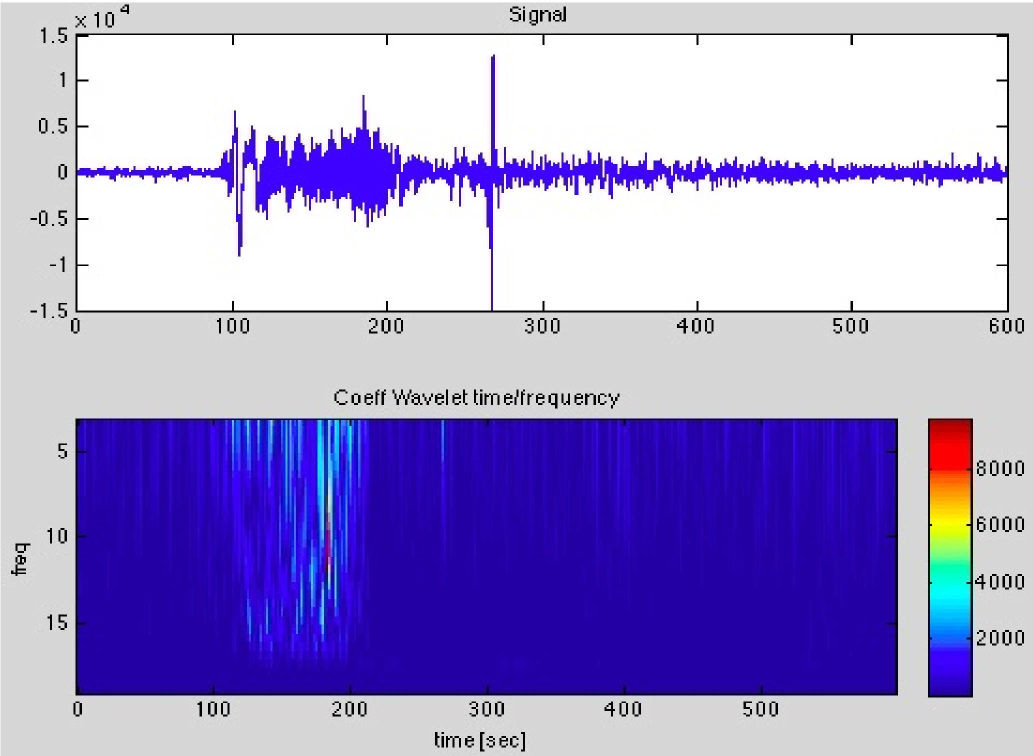


Figure 8.4: Wavelet Coefficients extracted from EEG signals

8.4 AR Modeling

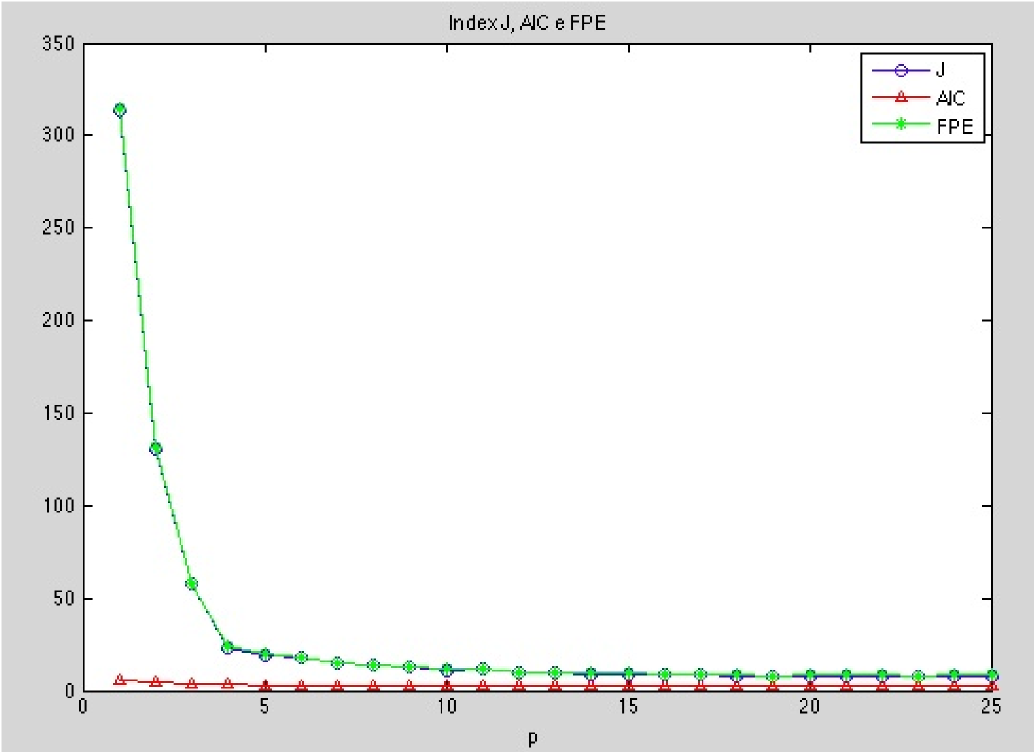
As already mentioned in the Chapter of Methods, a sampled EEG signal can be described as a function of its own past and a noise term:

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + e(n)$$

a_i are AR parameters and p is the order of the autoregressive model.
The optimal order p, for the signals considered, was established by implementing in the algorithm the following steps:

1. Identify the AR models of order p = 1:1:25 (using the command *ar*)
2. Calculate the values of **J**, **AIC** and **FPE** for each identification

The trends of the three indices to vary p were represented on the same graph:



For the dataset considered in this project, the optimal order according to index J is 10, while according to AIC is 19.
The order chosen for the EEG signals analyzed is order 10.

AR Validation

After choosing the model and the optimal order, it is a mandatory to test whether the model is able to represent the mechanism of data generation through the Anderson Test.
With the chosen order, the values outside the threshold, for the signals of the dataset, are 2,729% and the null hypothesis is accepted.

Figure 8.5 shows a variation of AR coefficients during the epileptic seizure in a representative study, so it is reasonable think to use these coefficients as inputs to the artificial neural network.

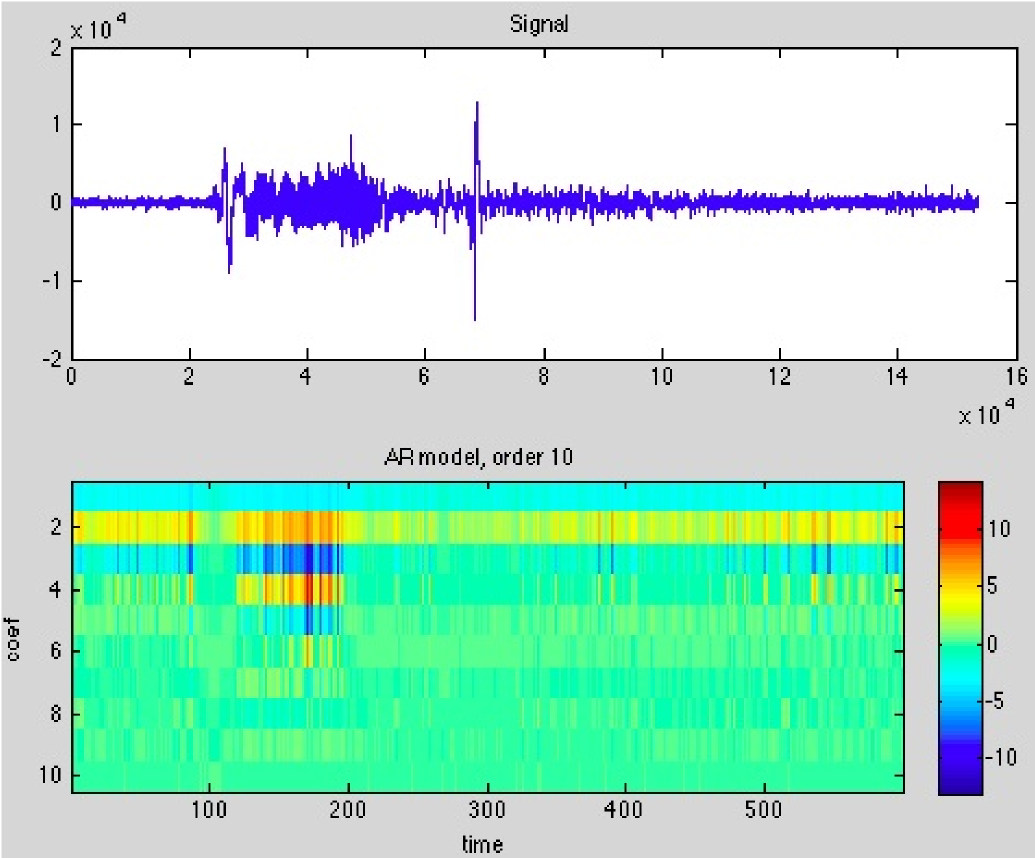


Figure 8.5: AR coefficients extracted from an EEG signal

8.5 Classifier

PCA

After the extraction of characteristics for each signal belonging to the database we will have available:

- for the wavelet method: a matrix composed of 54 rows (number of coefficients extracted for each epoch) and a number of columns equal to the epochs of the signal;

- for the AR method: a matrix consisting of 10 rows (number of coefficients needed to model each epoch) + 1 row containing the average amplitude of the signal for every epoch and a number of columns equal to the number of epochs;

The central idea of Principal Component Analysis (PCA) is to reduce the dimensionality of a dataset consisting of a large number of interrelated variables while retaining as much variation as possible.
This is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

Before using the features as input (*ptrain*) of the neural network, therefore we perform the following transformation:

$$[pn, ps1] = mapstd(ptrain);$$

$$[ptrans, ps2] = processpca(pn, 0.02);$$

The input vectors are first normalized using *mapstd*, in order to have zero mean and unit variance. In the example described in the second argument passed to the function *processpca* is 0.02.
This means that the function removes all the components of the normalized input vector, which contribute less than 2% of the total variation in the data set.

The *ptrans* matrix contains all the values so normalized, *ps*2 instead it is the structure that contains the transformation matrix of the principal components.

Each new input to the trained network will have to follow the same routine of normalization applied to the training set.

The Principal Component Analysis reduces the space of features to 22 lines, with regard to the parameters wavelet.
As regards the coefficients of the AR model, it was decided to maintain the matrix in input to the network of constant size (12 rows) because for each processed signal the PCA reduces the size of a variable number of features (for example for a signal deletes a row, while another eliminates three).
The neural network is trained with a matrix characteristics relative to a signal at a time, but requires training for each of the same a constant number of information input. To resolve this problem, holding constant the number of rows of the matrix to 12 and in the moment in which the PCA reduces the features from 12 to M (with M<12), 12-M rows are placed at the constant value 1, since the network automatically preprocesses inputs with

removeconstantrows.

Neural Network

Using Matlab functions is easy to create the neural network, is a little less easy to find the optimal number of nodes for this classification problem.
For this purpose it is necessary to resort to empirical test, varying the number of nodes in the network and validating the results. Once the results are obtained by testing various neural networks, it is chosen the network with the fewest number of False Positives and False Negatives, ie the one that has maximum Sensitivity and Specificity.

Fixed the number of nodes n, the functions to create the network, with the specifications explained in the chapter on Artificial Neural Network are:

```
HiddenLayer1=n;
net=feedforwardnet(HiddenLayer1);
init(net);

net.trainParam.epochs = 10000; %epochs number
net.trainParam.goal=1e-4;  % final error value
net.layers{1}.transferFcn='tansig'; %transfer function hidden layer
net.layers{2}.transferFcn='tansig'; %transfer function output layer
net.trainFcn='trainrp'; %training function
```

To train the network:

```
[net,tr]=train(net,ptrain,ttrain);
```

where the matrix *ptrain* contains the characteristics of each epoch and the vector *ttrain* is the expected output for every epoch, that output is given by the marker of the expert on the EEG track.

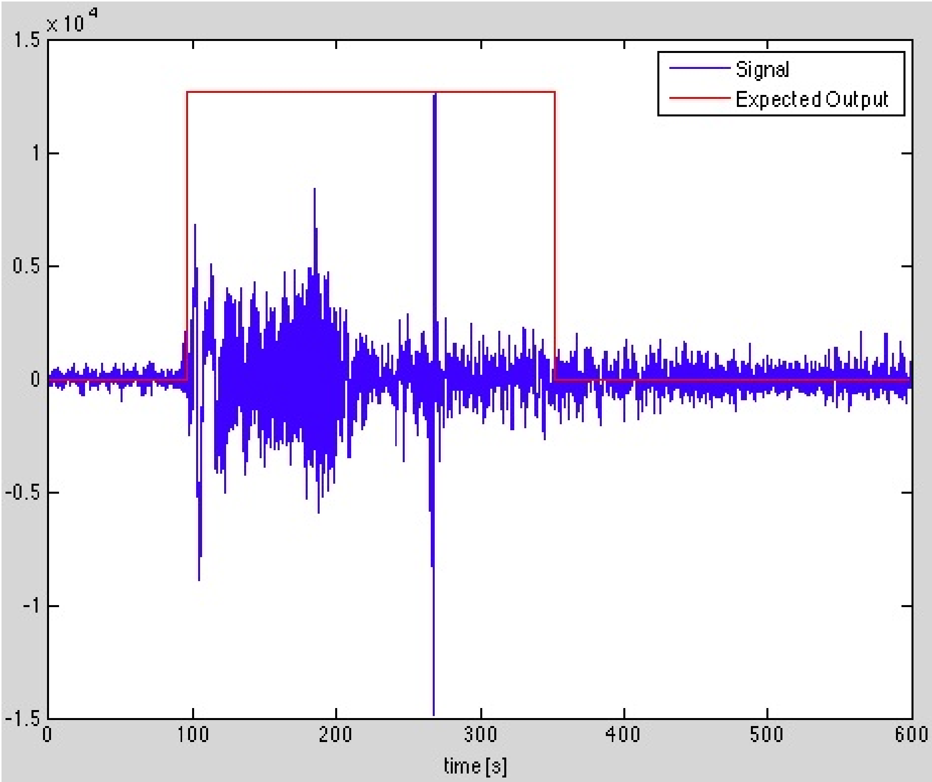


Figure 8.6: Signal EEG and its expected output (*ttrain*)

Validation

The training is repeated for all signals belonging to the database, and as already mentioned several times using the Leave One Out Validation, that is the training of network with all signals except one and then test it with the unknown signal, using the function:

```
output=net(ptrain);
```

After finding the output of the neural network on the unknown signal, Sensitivity is calculated as the ratio between the number of epochs properly classified as positive (seizure) and the total number of seizure epochs, and specificity as the ratio between the number epochs correctly classified as negative (noseizure) and the total of non-seizure epochs.

8.6 Expert System

The expert system to be used for the classification of different types of epilepsy must have a reliable basis of medical knowledge, on which are accumulated deductive rules and procedures for which the system is serving.
Once all the rules have been established, the program deals with the application in practical concepts contained in the data base and to give an answer to the problems proposed.

In this thesis, the expert system has been used to improve the reliability of the output, in particular, rules are applied to exclude crisis lasting less than four seconds, deemed as artifacts.

In addition, it is considered a single attack if the length of time that passes between a crisis and another is less than 2 seconds.

For classification by neural network, we used the features extracted from a single EEG channel, once calculated the output of the network for each channel it is possible to insert such responses in the expert system to exclude positive response in the case in which the crisis is detected only in a channel.

This expert system uses rules to confirm or exclude seizure, but other important uses can be implemented, such as those proposed in the next session devoted to future developments.

Summarizing, the rules applied in the algorithm to the output of the neural network are:

R1: IF length(seizure)<4 seconds, THEN seizure is discarded.

R2: IF time between two near seizures<2 seconds, THEN the two crises are treated as one.

Considering the output of network for all channel:

R3: IF seizure is in only one channel, THEN seizure is discarded.

Chapter 9

Results

9.1 Summary of Algorithm Parameters

A summary of the parameters used for seizure detection algorithm is given:

Input	Data Type	sEEG
	Sampling frequency	256 Hz
	Num Electrodes	19
	Montage	unipolar
Segmentation	Epoch Length	2 sec
	Overlap	50%
	Window	Rectangular
Feature Extraction	Wavelet Coef	'Morlet' (scale 11:64)
	AR Coef	order 10
Neural Network	Type	Feedforward
	Transfer Function	Tansig
	Data Division	Random
	Training Algorithm	Resilient Backpropagation
	Training Performance	Mean Squared Error
	Preprocess	PCA 2%

In the following tables the numerical results obtained in the validation process by varying the number of nodes and layers are presented.

9.2 Input: Wavelet Coefficients

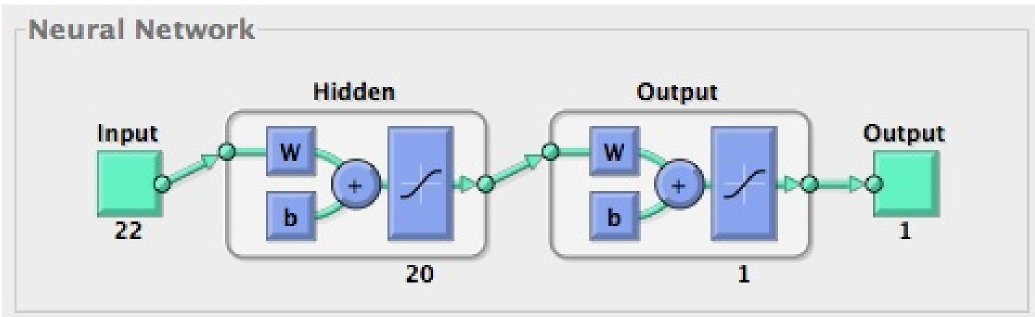
Hidden Layer	**Nodes**	**% Sensitivity**	**% Specificity**
1	3	45.1687%	87.36%
1	4	62.44%	63.07%
1	5	65.23%	77.43%
1	6	59.99%	85.61%
1	7	30.56%	91.15%
1	8	62.87%	70.66%
1	9	71.19%	89.53%
1	10	40.82%	70.71%
1	11	52.76%	83.31%
1	12	69.12%	70.34%
1	13	78.82%	90.17%
1	14	82.22%	86.75%
1	15	85.92%	93.56%
1	**20**	**94.46±7.15%**	**92.52±5.89%**
1	22	24.89%	62.67%
1	30	28.78%	93.69%
1	40	36.07%	67.72%
1	48	57.20%	81.12%
1	49	47.62%	85.31%
1	50	39.72%	72.01%
1	58	42.68%	66.36%
1	60	25.63%	51.86%
1	80	15.48%	59.88%

Hidden Layer	Nodes	% Sensitivity	% Specificity
2	5 3	51.35%	72.94%
2	5 5	66.49%	65.46%
2	10 5	71.63%	89.53%
2	10 10	78.64%	83.33%
2	20 20	49.70%	91.19%
2	40 40	34.80%	88.21%

Empirically, the best performance of the neural network is obtained with only one hidden layer consisting of 20 nodes.

1 Hidden Layer 20 nodes

The following figure shows visually the network:



With the neural network composed of such number of nodes, the results obtained for each patient considered in the database are specified in the following table

Patient n.	**%Sensitivity**	**%Specificity**
2	99.36	90.55
6	98.15	89.01
12	99.89	90.22
15	92.44	100
18	86.22	100
20	94.70	81.18
22	93.11	91.73
23	81.14	99.64
32	90.99	100
40	97.77	79.99
46	97.12	92.22
48	95.63	97.61
49	100	90.90
50	100	86.93
51	64.99	100
52	89.11	94.30
53	97.54	92.58
54	94.99	100
56	99.99	93.46
60	100	89.13
62	97.38	83.42
66	100	92.97
70	98.75	97.96
71	96.11	89.53
79	95.14	92.00
81	100	83.16
82	91.17	99.03
83	94.01	89.13
85	93.23	91.18
87	93.99	97.88
Mean ± SD	**94.46±7.15%**	**92.52±5.89%**

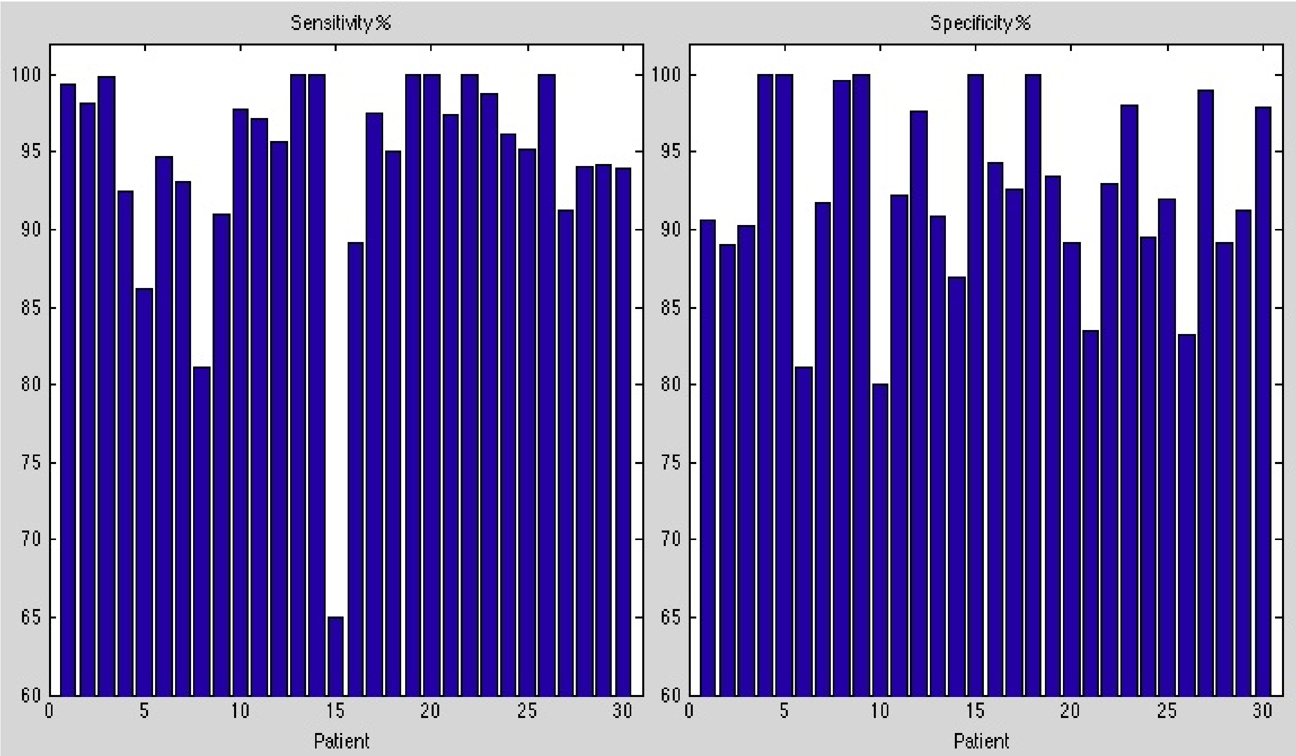


Figure 9.1: Results Sensitivity and Specificity obtained for each patient with an ANN classifier composed by 1 hidden layer of 20 nodes and input Wavelet coefficients

9.3 input: AR Coefficients + Amplitude

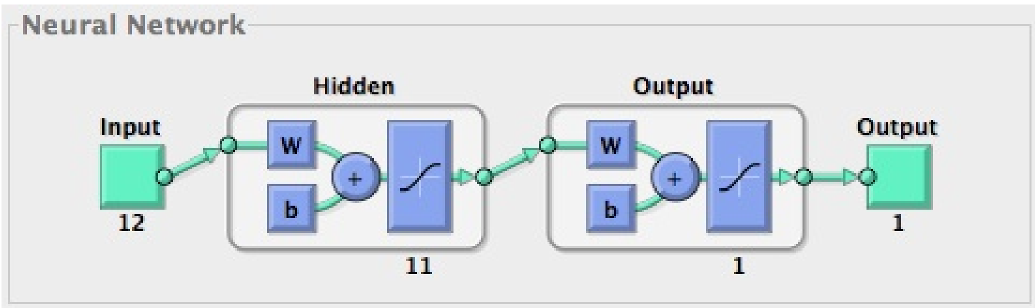
Hidden Layer	**Nodes**	**% Sensitivity**	**% Specificity**
1	3	78.45%	88.91%
1	4	38.58%	91.33%
1	5	50.93%	77.55%
1	6	79.96%	99.07%
1	7	68.19%	49.41%
1	8	56.91%	83.23%
1	9	71.73%	89.13%
1	10	56.25%	64.77%
1	**11**	**96.13±5.00%**	**97.46±5.43%**
1	12	94.60%	90.38%
1	13	77.75%	95.16%
1	14	72.42%	69.98%
1	15	95.80%	90.29%
1	20	71.02%	76.51%
1	24	95.85%	98.44%
1	30	47.95%	61.45%
1	40	39.03%	80.05%
1	48	24.43%	64.93%
1	49	54.83%	87.65%
1	50	47.83%	76.42%
1	58	31.85%	85.48%

Hidden Layer	Nodes	% Sensitivity	% Specificity
2	3 3	53.87%	89.01%
2	5 3	60.22%	68.31%
2	5 5	71.52%	83.68%
2	10 10	64.46%	70.44%
2	12 6	81.64%	80.97%
2	12 12	72.55%	87.77%
2	15 10	58.99%	92.28%
2	15 15	72.02%	78.62%
2	20 20	52.82%	81.13%
2	40 40	49.10%	69.68%

The best performance of the neural network, having as input the coefficients of the AR model and the average amplitude in the epoch, is obtained with only 1 hidden layer consists of 12 nodes.

1 Hidden Layer 11 nodes

The following figure shows visually this network:



With the neural network composed of such number of nodes, the results obtained for each patient considered in the database are specified in the following table

Patient n.	**%Sensitivity**	**%Specificity**
2	100	99.8
6	100	100
12	95.88	98.99
15	96.66	100
18	89.84	100
20	98.17	91.66
22	95.19	98.15
23	79.81	100
32	100	100
40	99.79	87.62
46	98.56	95.99
48	95.01	100
49	96.14	99.15
50	100	100
51	83.89	89.94
52	92.73	77.62
53	100	98.77
54	97.08	98.01
56	96.42	93.15
60	100	91.50
62	99.15	83.42
66	95.68	100
70	99.95	100
71	91.13	99.71
79	89.03	86.99
81	98.86	100
82	96.71	90.92
83	98.13	100
85	100	97.92
87	100	98.03
Mean±SD	**96.13 ± 5.00%**	**97.46±5.43 %**

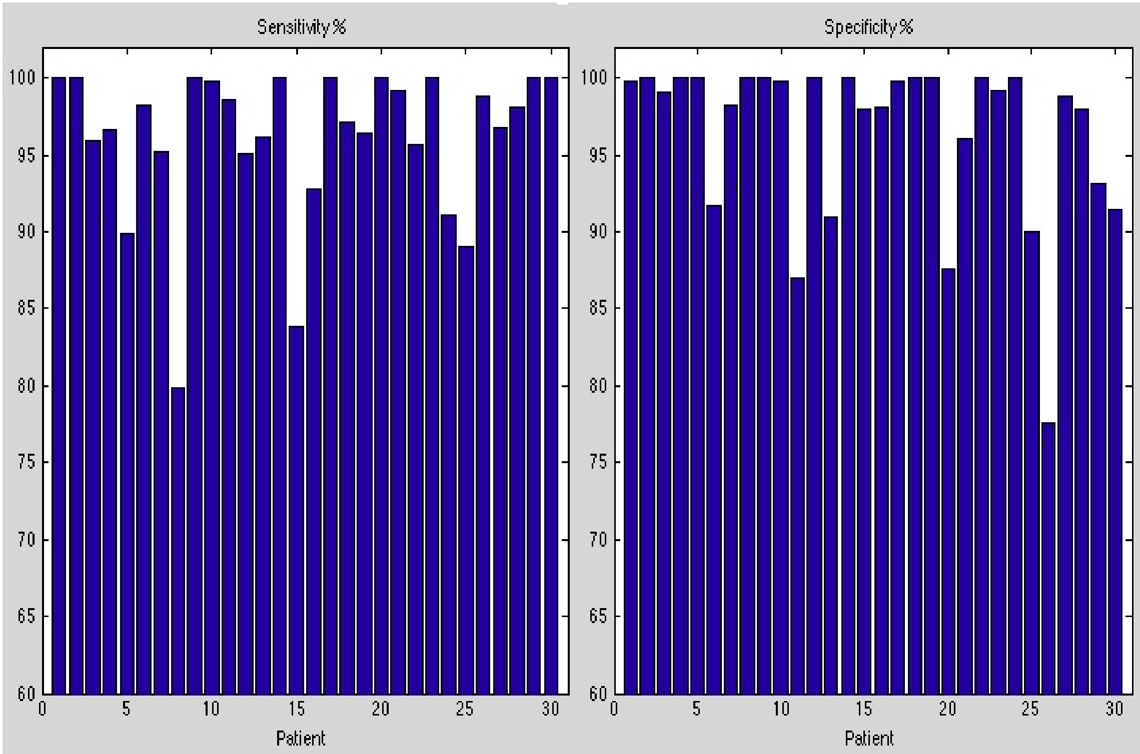


Figure 9.2: Results Sensitivity and Specificity obtained for each patient with an ANN classifier composed by 1 hidden layer of 11 nodes and input AR coefficients+amplitude.

The results obtained clearly show a performance improvement obtained with the AR coefficients, with respect to the network that uses the coefficients obtained through the continuous wavelet transform.

The simulation result indicated that AR model parameters robustly capture the abnormal brain activity and the proposed method improves the accuracy of the seizure detection.

To give more information to the network, since more information is given to the network in input more the network is able to correctly discriminate, it was decided to give both features as input. The following table presents the results obtained by having as input the coefficients obtained by wavelet transform and those obtained from the AR model.

9.4 input: Wav + AR Coefficients

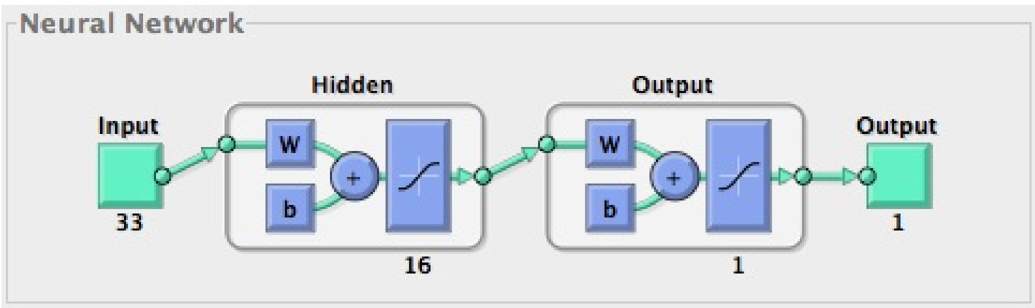
Hidden Layer	Nodes	% Sensitivity	% Specificity
1	3	88.09%	95.37%
1	4	64.92%	50.13%
1	5	74.37%	91.14%
1	6	67.60%	98.66%
1	7	81.79%	97.68%
1	8	60.44%	64.31%
1	9	98.25%	86.13%
1	10	83.17%	56.41%
1	11	73.21%	49.91%
1	12	97.90%	86.30%
1	13	81.43%	79.92%
1	14	65.27%	50.13%
1	15	83.52%	98.59%
1	**16**	**98.42±3.04%**	**94.08±4.46%**
1	17	75.71%	81.15%
1	18	63.34%	82.57%
1	19	39.92%	96.12%
1	20	42.15%	97.99%
1	25	70.58%	70.16%
1	30	86.67%	99.39%
1	34	76.89%	96.11%
1	40	97.20%	85.06%
1	45	57.03%	71.41%
1	50	58.81%	77.14%
1	68	48.98%	68.25%

Hidden Layer	**Nodes**	**% Sensitivity**	**% Specificity**
2	5 5	82.54%	82.97%
2	10 5	72.21%	85.79%
2	10 10	70.58%	94.39%
2	20 10	47.08%	98.04%
2	20 20	52.01%	88.82%
2	30 5	39.87%	91.11%
2	30 10	49.53%	70.24%
2	30 15	68.69%	91.43%
2	30 20	77.48%	73.38%
2	30 30	46.23%	93.66%
2	40 40	69.92%	96.35%

These results confirm the validity of this approach and it is immediate to conclude that the use of all features provides better results.

1 Hidden Layer 16 nodes

The optimal number of nodes for the network is shown in the following figure:



With the neural network composed of such number of nodes, the results obtained for each patient considered in the database are specified in the following table

Patient n.	**%Sensitivity**	**%Specificity**
2	100	98.60
6	100	99.17
12	98.68	90.01
15	99.87	89.15
18	99.15	99.68
20	100	95.27
22	99.79	89.32
23	87.16	95.99
32	98.85	99.16
40	99.01	100
46	100	86.15
48	97.19	91.06
49	91.63	100
50	100	97.74
51	91.73	90.99
52	95.99	100
53	100	96.00
54	100	94.63
56	99.51	97.15
60	100	93.81
62	100	85.18
66	98.85	100
70	100	91.22
71	97.66	95.91
79	98.75	93.21
81	100	90.90
82	98.79	92.12
83	100	90.18
85	100	89.55
87	100	90.36
Mean ± SD	**98.42±3.04%**	**94.08±4.46%**

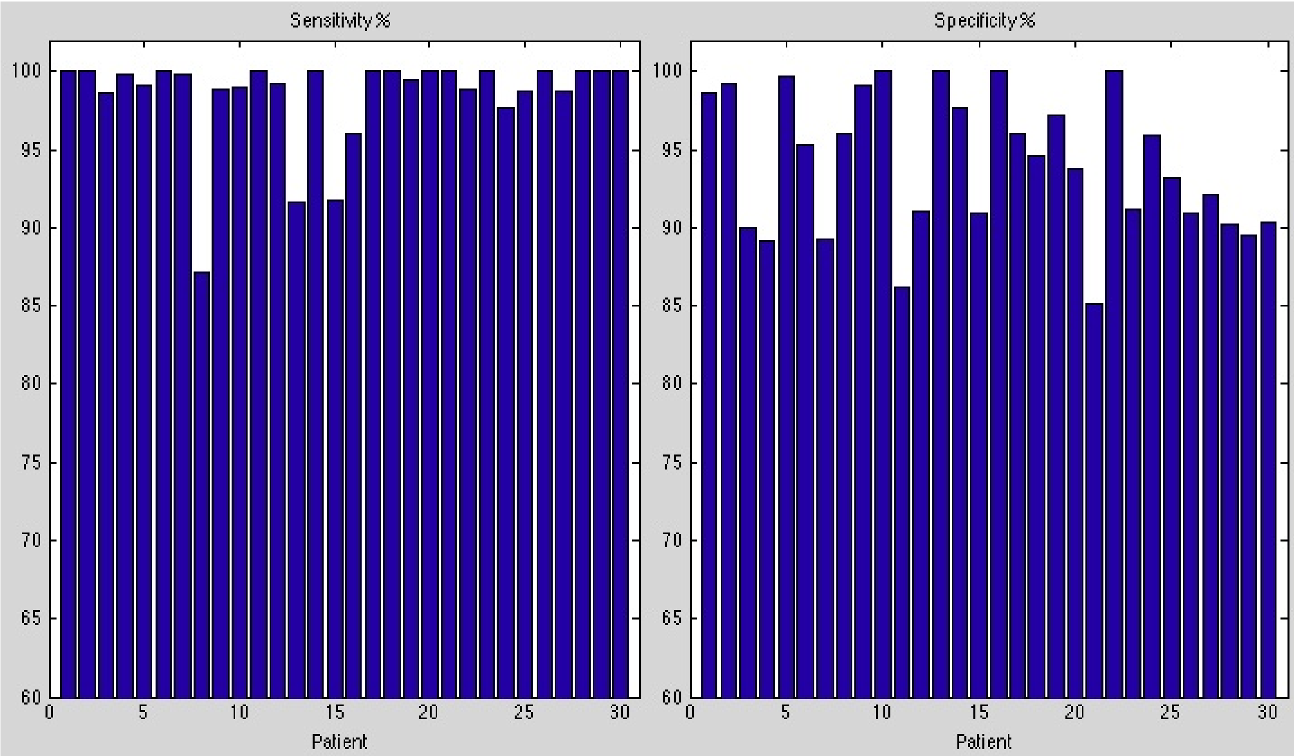


Figure 9.3: Results Sensitivity and Specificity obtained for each patient with an ANN classifier composed by 1 hidden layer of 16 nodes and input Wav coefficients+AR coefficients+amplitude.

Below we present an example of detection of the crisis by the EEG signal of a patient, after having trained the neural network with all signals except this.

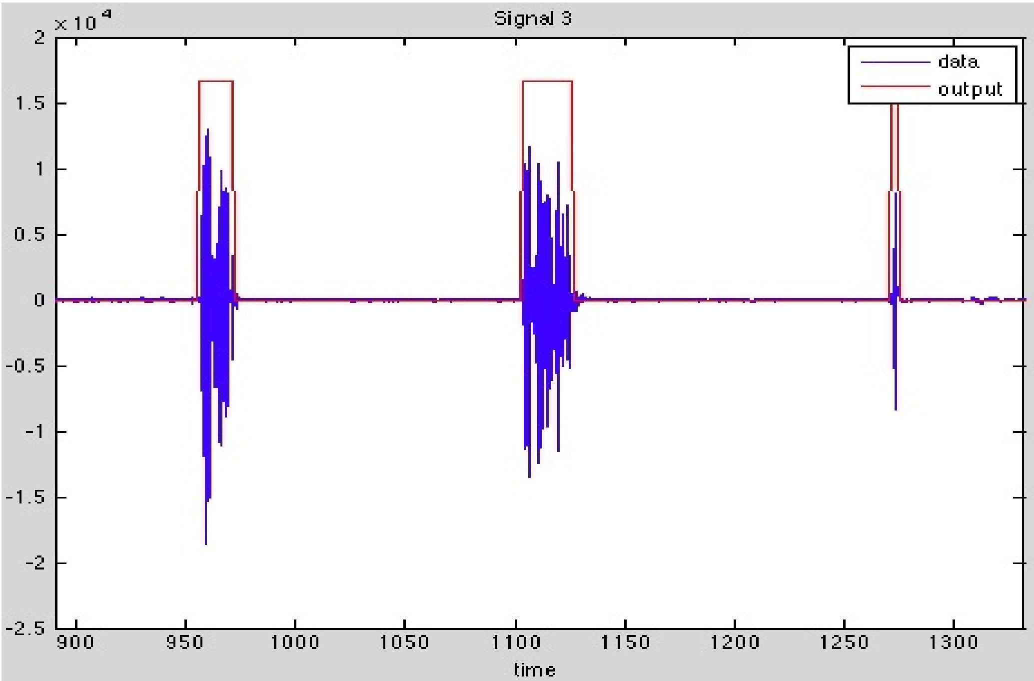


Figure 9.4: Result of seizure detection from EEG using as input Wav + AR Coefficients and Neural Network with 1 hidden layer consists of 16 nodes.

Chapter 10

Discussion

The purpose of this thesis project was to analyze and compare two methods for the automatic detection of seizures, in order to allow neurologists or specialists to save a lot of time for the interpretation of EEG tracks in long-term monitoring.

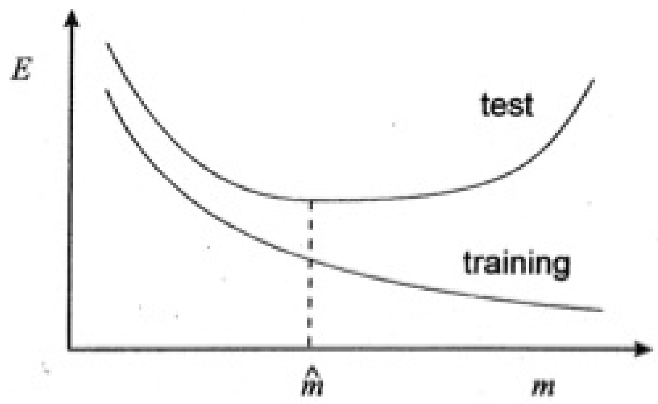
For this purpose two algorithms have been developed in parallel: one based on the features extraction from the EEG signal by continuous wavelet transform, the other describing the signal using autoregressive models.
Both the extracted features have proven to be a valuable tool for describing the signal, because during the crisis there is a change in the coefficients.

Artificial Neural Networks have been applied as a classifier and, in the chapter of results, several empirical attempts have been shown to determine the parameters of these networks in order to represent, with reasonable accuracy, the training data. The most important issue is to get a good performance in response to inputs that are not part of the training set.

This problem of "generalization" is, in practice, the choice of the optimal number of hidden units in order to:
(a) capturing the tendency of the training data;
(b) to have reliable answers when presented new data input (ie other than training).

As we have seen from the results, depending on the number of nodes m changes the error function E as a function of m. We would expect to obtain an error function E monotonically decreasing, as the addition of nodes should not increase the error (see figure). We present, then, the test data to several trained networks, and evaluate the corresponding value of E. The result

is that initially decreases as a function of m; However, when m reaches a high enough value, the phenomenon of overfitting enters in the game and the function begins to grow.
The optimal value of the number of nodes m is one that minimizes the error function test, and is given by $m = \hat{m}$ in the figure.



For each different network input we have found that the optimal value of $\hat{m}$ can feel satisfied of the results obtained through the use of neural network.

Pros:

* Neural networks work in parallel and are therefore able to treat many data while in traditional computers each data is processed individually and in succession. It is essentially a sophisticated statistical comes with a good noise immunity if some units of the system should malfunction, the network as a whole would have to reductions in performance but difficult would face a system crash.

* The development of the neural networks is distributed over many elements, ie there are many neurons that deal with the same operation.

* Neural networks must not be programmed to perform a task, but learn individually based on experience or with the help of an external agent.

Cons:

* The models produced by neural networks, although very efficient, can not be explained in the symbolic language of the human being: the results must be accepted "as they are", from which the definition of neural networks as **black box** systems.

* As with any modeling algorithm, even neural networks are efficient only for certain types of variables, and these must be chosen with care.

* Require a training phase of the system which optimizes the weights and this phase can take a long time if the number of training samples and of the analyzed variables is very large.

* There are no theorems or models to define the optimal network, so the success of a network depends heavily on the experience of the creator.

Leave One Out Cross Validation has allowed not only an evaluation of the algorithm performance and a comparison of the two methods implemented, but also to set the optimal parameters for the neural network.

The results obtained using a database of 30 subjects suffering from epileptic seizures has allowed us to show a better reliability for the network having in input the parameters of the AR model, compared to the wavelet coefficients but by giving input to the network both the parameters found with the AR model and the wavelet coefficients, the network is able to discriminate better.

Therefore we can conclude that the neural network, when it is used with the right number of hidden layer, the optimal number of nodes and with appropriate input information, can allow the detection of epileptic seizures and the results encourage the study of many aspects not covered in this project.

10.1 Guidelines for future development

There are still open questions about how best to exploit the potential of these new architectures that, as you can tell by neural networks in general, are very attractive, especially for their versatility and ability to solve problems of various kinds.

The results of this thesis also encourage the deepening of the topics with the aim to encourage and support patients in the detection of epileptic seizures. In this regard we now present some open problems that would be interesting to investigate in the future:

a) Use of all EEG channels: Here, the network information derives from a single EEG channel and use the outputs obtained in the different channels for the discrimination of the crisis through expert system. Another reasonable hypothesis, which would be interesting to investigate, was put directly in the input to the network information relating to all the EEG channels.

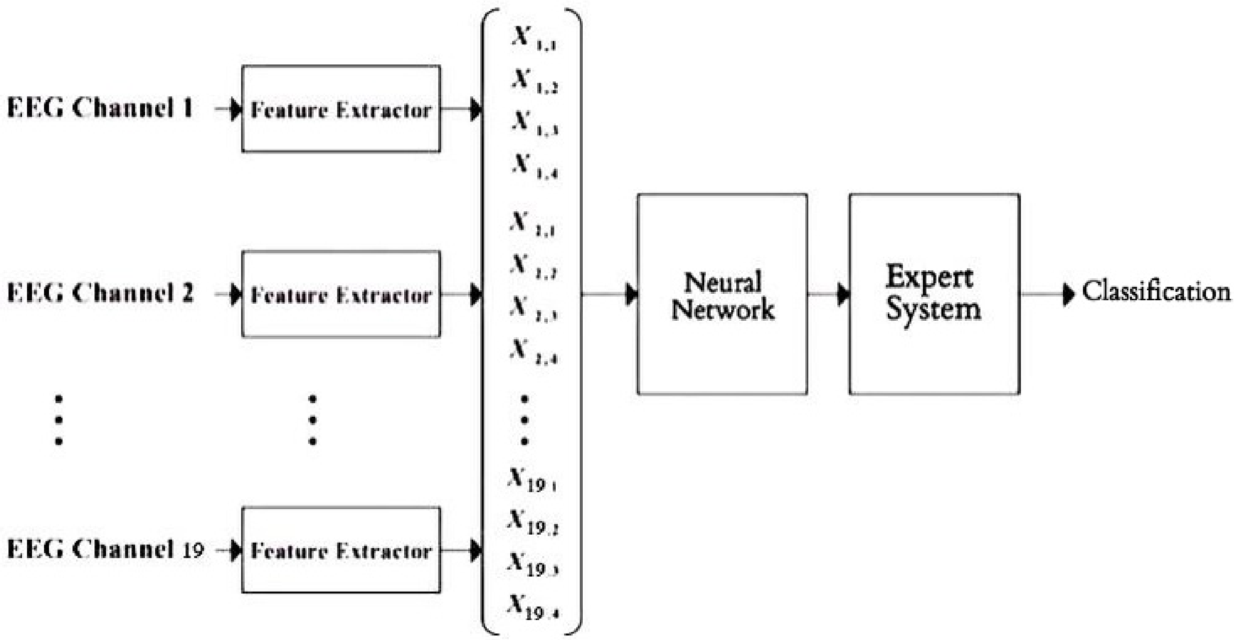


Figure 10.1: Neural Network with information input of all EEG channel

b) Classification of different types of epilepsy: Another important aspect that is of great interest concerns the classification of the various types of epilepsy from the response of the various EEG channels. To carry out this task is essential the consultation with a specialist, but to make it active is enough properly implement these rules with an expert system or by using the neural network.

c) Identification of new useful features to describe the signal: Neural networks used in this project are able to detect the presence of seizures in the EEG, to increase the accuracy of such detection may be useful to identify other characteristics related to the signal that discriminate changes in the path due to the crisis from those due to artifacts.

d) Online Detection: use of the algorithm for the prediction of epileptic seizures from the past informations of the EEG signal.

Appendix A

Process of Software Certification

The purpose of this chapter is to provide guidance to the application of all the provisions which apply to software within the scope of the Medical Device Directives.
Medical devices contribute increasingly, thanks to continuous technological innovations to improve the quality of care and to extend the number of people who can benefit.
Also on the regulatory and administrative requirements, there has been a rapid evolution.

Medical Device: any instrument, apparatus, appliance, *software*, material or other, whether used alone or in combination, including the software necessary for its proper application intended by the manufacturer to be used for human beings for the purpose of:

- Diagnosis, prevention, monitoring, treatment or alleviation of disease;
- Diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap;
- Investigation, replacement or modification of the anatomy or of a physiological process;
- Control of conception.

A.1 Software

Software is an active Medical Device and it could be:

A software that is a medical device or is an accessory of a medical device (**SUBJECT TO CE MARKING**), for example for the analysis of cardiac signals of long duration from Holter or for the calculation, estimation, modeling of surgical placements or software for the comparative long-term monitoring of recorded images for diagnostic oncology.

A software that is an integral part or component of a medical device (**NOT SUBJECT TO CE MARKING**), e.g. an integrated software that controls the operation of a infusion pump.

Software that is not covered by the Medical Device Directive, for example software used for the administrative management of patients, software used for the training of medical personnel or software used for the maintenance of medical devices or components.

A.2 Procedures for certification

Procedures for certification include modules for certified to ensure the quality system, particular attention should be paid to the *design control*, *control of documents and records of quality*, the *responsibilities of the management* and to *control of combinations*.

Design control: states that the development cycle for the product should be defined on the basis of planning, risk management, verification and validation (ref. ISO 13485).

Control of documents and records of quality: must be specified the procedures for document control and configuration management.

Responsibilities of the management: must provide trained personnel and define the different responsibilities in terms of risk analysis, verification and validation.

Control of combinations: in the case of combined use with other devices must always consider the aspects of safety and performance.

A.3 CE Certification

The verification should involve both a review of records related to the cycle of development both an appropriate inspections and tests regarding the verification of system software configuration management and program checksum. The review must always ensure repeatable performance and effectiveness of the software.

If the level of complexity of a software is such that the execution only of the final test is not able to ensure the detection of systematic errors. And 'essential to check the management procedures, therefore this kind of procedure is not considered adequate for the CE marking.

A.4 Quality assurance of the production

In order to make adequate guarantee the quality of production must necessarily take into consideration:

- If you have defined a method of development which includes the concepts of the development cycle, taking into account the aspects of planning, risk management, verification and validation
- If the procedures relating to the control of documentation and configuration management have been applied
- When you have been trained and have defined the various responsibilities in terms of risk analysis, verification and validation
- If the system owner and the supporting software have been validated for the particular application

The quality assurance of the production is not be appropriate in cases where the final inspections and testing are not able to guarantee an adequate level of safety due to the inability to identify any systematic errors.

A.5 Technical File:

The Manufacturer must necessarily constitute the **Technical File**, in the same way as all the other Medical Devices.
As regards the positioning of the CE, in the case where the identifier of the software is on the screen, you can place the marking CE in its vicinity, if the software is sold via CD, the CD must be properly marked. In any case, the CE marking must be present on the accompanying documentation.

The **Clinical Efficacy** must be shown of the performance of the software and its clinical effectiveness, in accordance with the intended use as declared by the manufacturer. Must be defined and documented testing protocols.

The **Test Reports** must use harmonized standards and standards established by the state of the art.

For devices which incorporate software or which are medical software in themselves, the software must be validated according to the state of the art taking into account the principals of development lifecycle, risk management, validation and verification. [29]

A.6 State of the Art:

- **CEI EN 60601-1-4:1997 + A1:2000** Medical electrical equipment Part 1: General requirements for safety
  - Collateral standard: Programmable electromedical systems
  The Collateral standard applies to the safety of equipment and electromedical systems that incorporate programmable Electronic Subsystems (PESS) also called Programmable Electromedical Systems (PEMS).
  It specifies the requirements for the project manufacturing.
  It refers to the system architecture, detailed design and implementation, including the development of the software, any changes made, verification and validation, and finally the marking and documentation attached.
  The standard does NOT apply to the implementation of the hardware, the reproduction of the software, installation, and commissioning and operation and maintenance

- **UNI EN ISO 9001:2000** Systems of quality management or **UNI EN ISO 13485:2004** Quality management systems - Requirements for regulatory purposes

The rule is also about the software used for the production
It stresses the need to ensure the identification and traceability of the software.

- **IEC 14971:2007** - Application of risk management to medical devices
- **IEC 62304**, Ed. 1: Medical device software - Software life cycle processes
  This standard is applied to the development and maintenance of medical software and when the software itself is a medical device either when the software is an integral part of the medical device. It does not cover the validation and release of the software.
- **FDA guidelines**
  Guidance for the content of Premarket Submissions for Software contained in Medical Devices: 2005
  General Principles of Software Validation; Final Guidance for Industry and FDA Staff: 2002

In summary:
- Define whether the DM software is based on its intended use
- Classification according to the rules of the DM active
- Annex which provides for the management of software quality (Annex II or V)
- Technical documents with the use of technical standards currently in force (IEC 62304, ISO 14971, 60601-1 Series)

Bibliography

[1] Proposal for revised clinical and electroencephalographic classification of epileptic seizures. From the Commission on Classification and Terminology of the International League Against Epilepsy, *Epilepsia*, 22(4):489-501, 1981.

[2] A.T. Berg, S.F. Berkovic, M.J. Brodie, J. Buchhalter, J.H. Cross, W. van Emde Boas, Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, *Epilepsia*, 51(4):676-685, February 26th 2010.

[3] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press Oxford, 1995

[4] Edward H. Reynolds, *Introduction to Epilepsy: History of Epilepsy*, *Cambridge University Press*, pp. 1-6, 2012.

[5] J. Engel, *Seizure and Epilepsy*, *Oxford University Press*, 2013.

[6] R.S. Fisher, W. van Emde Boas, W. Blume, C. Elger, P. Genton, P. Lee and J. Engel, Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) , *Epilepsia*, 46(4):470-472, 2005.

[7] J. Gotman, and P. Gloor, *Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG*, Electroencephalography and Clinical Neurophysiology, vol. 41, pp. 513-529, 1976.

[8] J. Gotman, *Automatic recognition of epileptic seizures in the EEG*, Electroencephalography and Clinical Neurophysiology, vol. 54, no. 5, pp. 530540, 1982.

[9] J. Hughes. Did all those famous people really have epilepsy? *Epilepsy and Behavior*, 6(2):115-139, Mar 2005.

[10] Y. U. Khan, O. Farooq, and P. Sharma, *Automatic detection of seizure onset in pediatric EEG*, International Joural of Embeded Systems and Applications, vol. 2, no. 3, pp. 8189, 2012.

[11] A. Kharbouch, A. Shoeb, J. Guttag, and S. S. Cash, *An algorithm for seizure onset detection using intracranial EEG*, Epilepsy Behav., vol. 22, pp. S29-35, 2011.

[12] M. K. Kiymik, I. Guler, A. Dizibuyuk, and M. Akin, *Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application, Computers in Biology and Medicine*, vol. 35, no. 7, pp. 603 616, 2005.

[13] J. Malmivuo, R. Plonsey, Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields, *Oxford University Press*, New York, chapter 13, 1995.

[14] J. L. Perez Velazquez, Jeanne Zhen Huo, L. Garcia Dominguez, Y. Leshchenko, O. Carter Snead III, Typical versus Atypical Absence Seizures: Network Mechanisms of the Spread of Paroxysms, *Epilepsia*, 48(8):15851593, 2007.

[15] H. Qu, and J. Gotman, *A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device*, IEEE Trans. on Biomed. Engg., vol. 44, no. 2, pp. 115122, Feb 1997.

[16] J. Roger, M. Bureau, et al, Epileptic syndromes in infancy, childhood and adoles- cence, *John Libbey*, 2nd ed. London, England, pp. 106-312, 1985.

[17] M. E. Saab, and J. Gotman, *A system to detect the onset of epileptic seizures in scalp EEG*, Clinical Neurophysiology, vol. 16, no. 2, pp. 427-442, 2005.

[18] Seizures and epilepsy: Hope through research. National Institute of Neurological Disorders and Stroke, *NIH Publication*. March 22, 2011.

[19] B. Shneker and N. Fountain. Epilepsy. *Disease a Month*, 49(7):426-478, Jul 2003.

[20] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. Ted Treves, and J. Guttag, *Patient-specific seizure onset detection*, Epilepsy and Behavior, 5(4):483498, 2004.

[21] A. Shoeb, and J. Guttag, *Application of Machine Learning To Epileptic Seizure Detection*, in Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010, pp. 975- 982.

[22] T. L. Sorensen, U. L. Olsen, I. Conradsen, J. Henriksen, T. W. Kjaer, C. E. Thomsen, and H. B. D. Sorensen, *Automatic epileptic seizure onset detection using Matching Pursuit: A case study*, in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 2010, pp. 3277-3280.

[23] G.E. Polchronaki, P. Ktonas, S. Gatzonis, A. Siatouni, H. Tsekou, D. Sakas. K.S. Nikita, *A system to detect the onset of epileptic seizures in scalp EEG*, IEEE transaction, July 5th, 2008.

[24] O. Temkin, The Falling Sickness, *Johns Hopkins Press*, Baltimore, MD, page 250, 1971.

[25] H. Wang, Z. Chen, S. Zou, L. Zhao, *The continuous analysis of EEG's alpha wave by morlet wavelet transform*, 2010 Aug;27(4):746-8, 752.

[26] S. B. Wilson, M Scheuer, R Emerson, and J Gabor. Seizure detection: evaluation of the Reveal algorithm. Clin. Neurophysiol., 115(10):22802291, 2004.

[27] G. Zanchin. Considerations on "The Sacred Diseas" by Hippocrates. *Journal of the History of the Neurosciences*, 1(2):91-5, Apr 1992.

[28] http://www.mayoclinic.org/diseases-conditions/petit-mal-seizure/basics/treatment/

[29] http://www.aiic.it/documents/35265/80468/140012772Bianchi.pdf

[30] http://dali.feld.cvut.cz/ucebna/matlab/toolbox/wavelet/cwt.html

Ringraziamenti

Giunta al termine di questa esperienza alcuni ringraziamenti sono doverosi.
In primo luogo, un ringraziamento va alla prof. Toffolo per la sua professionalità e per avermi stimolata a migliorare costantemente.
Vorrei inoltre ringraziare il prof. Ruggeri per avermi dato l'opportunità di lavorare presso l'azienda Micromed S.p.a, che si occupa di argomenti che destano da sempre il mio interesse.

La mia esperienza di tesi in azienda ha contribuito alla mia crescita culturale e personale, vorrei ringraziare in particolare:
Cristiano Rizzo perchè senza di lui tutto questo non sarebbe stato possibile,
Raffaele Orsato per aver sopportato pazientemente tutti i miei infiniti dubbi, domande e curiosità sull'argomento,
Nicola Rizzo per la sua gentilezza e il suo ottimismo,
Gianpietro Favaro, che ha sempre trovato tempo per ascoltarmi e insegnarmi, per la sua enorme conoscenza nell'ambito ingegneristico e nella vita,
Monica Camillo e Chiara Borgogno per avermi dato tutto il supporto morale di cui avevo bisogno.
Sono contenta di non aver trovato solo "colleghi", ma persone vere che hanno sempre creduto nelle mie capacità, anche quando io ero la prima a dubitarne.

Il ringraziamento più grande va alla mia famiglia, che ha sostenuto in tutti i modi i miei studi e le mie scelte di vita, permettendomi di raggiungere il mio traguardo.
Grazie a Giorgio e al suo tacito supporto in questi anni.

Un grazie va anche a Loris per aver sempre creduto in me, sostenendomi e incoraggiandomi. Grazie per aver condiviso con me gioie, soddisfazioni, malumori e incertezze dal primo all'ultimo giorno ad Ingegneria.

Grazie alle mie compagne di corso Valentina e Michela che hanno reso speciale ogni mio giornata ingegneristica, per le ore di studio passate insieme e anche per le ore di risate con le lacrime agli occhi, sarà questo il ricordo che porterò sempre nel cuore di questi cinque anni. Sono fortunata ad avere due amiche come voi.

Grazie a tutti coloro che ci sono stati quando ne avevo bisogno e a coloro che avrebbero voluto esserci.

Ancora mille e mille grazie perchè senza di voi non avrei raggiunto questo importante traguardo.