People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA – Boumerdes

Institute of Electrical and Electronic Engineering

Department of Electronics

Final Year Project Report Presented in Partial Fulfilment of the Requirements for the Degree of

MASTER

In **Electronics**

Option: **Computer Engineering**

Title:

EEG Signal Feature Extraction and Classification for Epilepsy Detection

Presented by:

- **FALKOUN Noussaiba**
- **OUAKOUAK Ferial**

Supervisor:

Dr. CHERIFI Dalila

Registration Number:……/2020

Dedications

« Behind every young child who believed in himself is a parent who believed first » This humble work is dedicated to my amazing parents for their unconditional love and support throughout the years. And most of all, for always believing in me even when I failed to do so. I cannot imagine reaching this point in my life or even accomplishing this work without them around. This dedication is a token of my appreciation and the least I can do to thank them.

I would also like to thank my friend Ferial with whom I shared the whole experience of this work, the bad, good and funny moments. It has been an unforgettable journey!

Noussaiba

I dedicate this work to my beloved parents Ali and Fatma, who continually provide their moral, spiritual, emotional, and financial support during my educational career.

To my brothers Mohand, Lyes and Nouredine, my uncles, my aunts and my grandparents whom I am truly grateful for having them in my life.

To all my friends, thank you for your encouragement, especially Amina, Nadia, Rania and Yasmine with whom bad moments of these five years have been good, and the good ones unforgettable.

Special thanks to my friend Noussaiba and my best friend SAMIRA for their advice and patience.

Ferial

Acknowledgments

First of all, we would like to thank Allah for His abundant grace in getting us through this process. Then, we would like to express our sincere gratitude to our supervisor Dr. CHERIFI Dalila for providing her invaluable guidance, comments and suggestions throughout the course of this project. Last but not least, we would like to thank all the teachers and the staff of the Institute of Electrical and Electronic Engineering, for their great devotion towards their professions, and for providing a convenient working environment.

Abstract

Epilepsy is a neurological disorder of the central nervous system, characterized by sudden seizures caused by abnormal electrical discharges in the brain. Electroencephalogram (EEG) is the most common technique used for Epilepsy diagnosis. Generally, it is done by the manual inspection of the EEG recordings of active seizure periods (ictal). Several techniques have been proposed throughout the years to automate this process.

In this study, we have used three different approaches to extract features from the filtered EEG signals. The first approach was to extract eight statistical features directly from the time-domain signal. In the second approach, we have used only the frequency domain information by applying the Discrete Cosine Transform (DCT) to the EEG signals. In the last approach, we have used a tool that combines both time and frequency domain information, which is the Discrete Wavelet Transform (DWT). Six different wavelet families have been tested with their different orders resulting in 37 wavelets. The extracted features are then fed to three different classifiers k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), and Artificial Neural Network (ANN) to perform two binary classification scenarios: healthy versus epileptic (mainly from interictal activity), and seizure-free versus ictal. We have used a benchmark database, the Bonn database, which consists of five different sets. In each scenario, we have taken different combinations of the available data. For Epilepsy detection (healthy vs epileptic), the first approach performed badly. Using the DCT improved the results, but the best accuracies were obtained with the DWT-based approach. For seizure detection, the three methods had a good performance. However, the third method had the best performance and was better than many state-of-the-art methods in terms of accuracy. After carrying out the experiments on the whole EEG signal, we separated the five rhythms and applied the DWT on them with the Daubechies7 (db7) wavelet for feature extraction. We have observed that close accuracies to those recorded before can be achieved with only the Delta rhythm in the first scenario (Epilepsy detection) and the Beta rhythm in the second scenario (seizure detection).

Keywords: *Electroencephalogram (EEG), Epilepsy, feature extraction, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), classification, k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), Artificial Neural Network (ANN).*

Table of contents

List of figures

List of tables

List of abbreviations

acc	Accuracy.
ANN	Artificial Neural Network.
bior	Biorthogonal wavelet.
coif	Coiflets.
CWT	Continuous Wavelet Transform.
db	Daubechies.
DCT	Discrete Cosine Transform.
DWT	Discrete Wavelet Transform.
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
fMRI	functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
FT	Fourier Transform
HPF	High-Pass Filter
IMF	Intrensic Mode Functions
IQR	Interquartile Range
Iz	Inion
k-NN	k-Nearest Neighbor
LPF	Low-Pass Filter.
LS-SVM	Least Square- Support Vector Machine.
LSTM	Long Short Term Memory
MLP	Multi Layer Perceptrons
MRI	Magnetic Resonance Imaging
Nz	Nasion
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PNN	Probabilistic Neural Networks
QP	Quadratic Programming
RBF	Radial-Basis Function
ReLU	Rectified Linear Unit
RNN-LSTM	Recurrent Neural Network- Long Short Term Memory
sen	Sensitivity
spe	Specificity
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
WHO	World Health Organization
WT	Wavelet Transform

General introduction

The human brain is the most complex and mysterious organ of the human body, consisting of billions of neurons. It is considered as an electro-chemical machine because neurons exploit chemical reactions to generate electrical signals. These electrical signals can be monitored through different scientific techniques such as Electroencephalography (EEG), Magnetic Resonance Imaging (MRI), functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET). EEG is the most used technique to capture brain signals due to its ease of use, its excellent resolution and its low cost. It is used in the medical environment more precisely in the diagnosis and treatment of mental and neurological disorders (Alzheimer, Dementia....) and more particularly in the case of Epilepsy.

According to an estimate of the World Health Organization (WHO), Epilepsy affects around 50 million people worldwide. Epilepsy is characterized by recurrent and sudden seizures. These seizures are the result of a transient and unexpected electrical disturbance of the brain and excessive neuronal discharge that is evident in EEG. The detection of epileptic seizures by visual scanning of a patient's EEG data is a tedious and time-consuming process. In addition, it requires an expert to analyze the entire length of the EEG recordings. Moreover, the diagnosis of Epilepsy is nearly impossible from the seizure-free EEG recordings. As a result, it is necessary to develop a robust and a reliable automatic classification and detection system for Epilepsy diagnosis. For this aim, several automated EEG signal classification methods, using different approaches, have been proposed. However, most of them deal with seizure detection only.

In this work, an analysis of EEG signal is performed to detect Epilepsy during both ictal and interictal states. This is executed using three different techniques of feature extraction and three distinct classification algorithms. In order to compare the performance of these methods, each algorithm is tested on a real dataset which consists of three subject groups: healthy subjects (normal EEG), epileptic subjects during a seizure-free interval (interictal EEG), and epileptic subjects during a seizure (ictal EEG).

This project report consists of four chapters. The **first chapter** aims to introduce the EEG signal and the Epilepsy. For this, a brief description of the anatomy and the activity of the brain is given followed by some basics of the EEG technique. The **second chapter** explains two steps of EEG signal analysis. The first step is the preprocessing and the second is feature

extraction where three techniques are described. Classification is a very important step of EEG signal analysis, for this the **third chapter** is devoted for machine learning and classification algorithms. Our choice was to use three classifiers, which are k-Nearest Neighbors, Support Vector Machine and Artificial Neural Network. Ultimately, **chapter four** presents the experimental results obtained from applying the algorithms described in chapter two and three on the Bonn dataset. Finally, conclusions of this work and possible perspectives are drawn followed by an appendix where extended experimental results are presented.

Chapter 1

Review about the Brain and Electroencephalogram

1.1 Introduction

This chapter gives basic knowledge about the brain and the electroencephalogram technique. Before diving into the main topic of our project, we saw fit to give a brief idea about the brain, since Epilepsy is a brain disorder, and give an insight about the EEG technique with some of its advantages and drawbacks and how it is used for Epilepsy diagnosis.

1.2 Brain Anatomy

The brain is one of the largest and most complex organs in the human body. It is the main organ of the human central nervous system (CNS). It functions as the coordinating center of sensation and intellectual and nervous activity. On a basic level, the brain can be divided into three major parts: cerebrum, cerebellum and brainstem, as shown in figure 1.1.

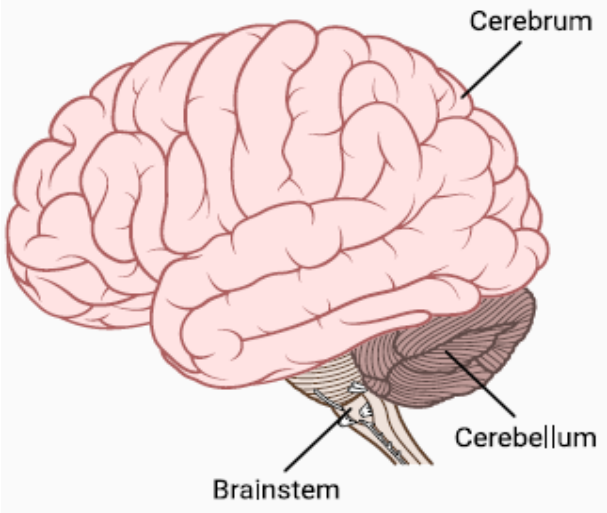


Figure 1.1 The main parts of the brain [1].

1.2.1 The Cerebrum

The Cerebrum or cortex is the forward-most portion and largest part of the human brain. It is composed of the right and left hemispheres, which are joined by mass of nerve cells called the corpus callosum. It is responsible of conscious thinking such as judgment, reasoning and learning, and sensory processing such as initiation and coordination of movement [1,2].

1.2.2 The Cerebellum

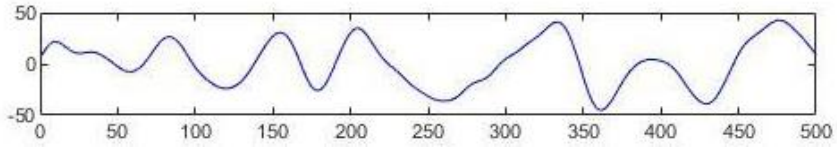
The cerebellum is located behind the top part of the brainstem. It has two hemispheres, which have highly folded surfaces. It receives information from the sensory systems, the spinal cord and some other parts of the brain to fine-tune motor activity [3]. The cerebellum contributes to regulation and control of fine movements, posture and balance [1].

1.2.3 The Brainstem

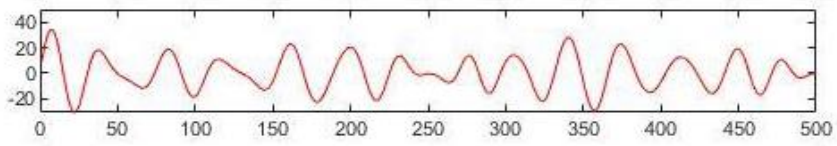
The brainstem is the area at the base of the brain that lies between the deep structures of the cerebral hemispheres and the cervical spinal cord [4]. It serves a critical role in regulating certain involuntary actions of the body such as heartbeat, breathing, bladder function and sense of equilibrium.

1.3 Brainwaves

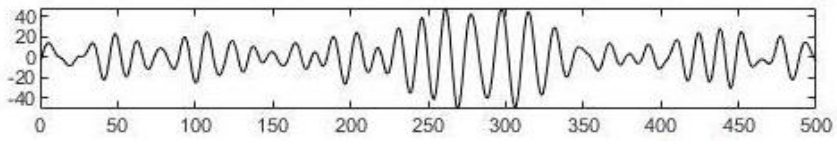
There are about 85 billion neurons in the human brain [1]. Each individual neuron connects to thousands of others. The communication between neurons is at the root of all the human thoughts, emotions and behaviors. It happens through small electrical currents that travel along the neurons and throughout enormous networks of brain circuits [5]. The synchronized electrical pulses resulting from this communication produce the brainwaves. Our emotions and activities select which brainwaves are dominant: the slow brainwaves when we feel tired and sleepy, the moderate brainwaves when we are calm and relaxed, or the fast brainwaves when we feel wired or hyper-alert. They are divided into five bands or rhythms according to the frequency: the Delta rhythm, the Theta rhythm, the Alpha rhythm, the Beta rhythm and the Gamma rhythm. The frequency boundaries of each band change from one source to another. In this research, we considered the frequency bands defined in Wikipedia. Figure 1.2 shows the five rhythms extracted from a 3 seconds (approximately) EEG recording of a healthy subject with eyes open. The amplitude is in microvolts.



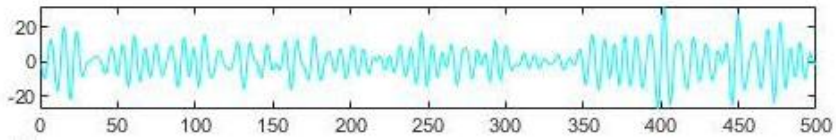
(a)



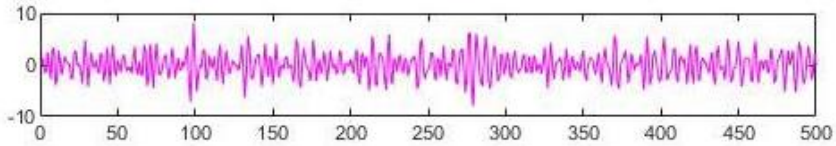
(b)



(c)



(d)



(e)

Figure 1.2 EEG recordings of the five rhythms (a) Delta (b) Theta (c) Alpha (d) Beta and (e) Gamma.

1.3.1 Delta Rhythm (<4Hz)

Delta waves are the slowest and highest amplitude brainwaves [1]. They are generated in deep meditation and dreamless sleep. Delta waves suspend external awareness and are the source of empathy. This is also the state where healing and regeneration are stimulated [6].

1.3.2 Theta Rhythm (4Hz-7Hz)

Theta waves occur most often in sleep and are strongly detectable when we are dreaming. They are also dominant in deep meditation. In theta, our senses are withdrawn from the external world and focused on signals originating from within. Research has also shown a positive association of theta waves with memory, creativity and psychological well-being [5,6].

1.3.3 Alpha Rhythm (8Hz-15Hz)

Alpha waves were the first to be discovered and they are easily observed [5]. They are stronger during physical relaxation with eyes closed and weaker during mental or physical activity with eyes open [1]. Alpha is the resting state for the brain. Its waves aid overall mental coordination, calmness, alertness, mind/body integration and learning [6].

1.3.4 Beta Rhythm (16Hz-31Hz)

Beta is a fast activity. It is present when we are alert and our attention is directed towards cognitive tasks and the outside world. It is dominant in the normal waking state of consciousness when we are busy thinking actively [6].

1.3.5 Gamma Rhythm (>32Hz)

At the moment, Gamma frequencies are the black holes of EEG research as it is still unclear where they are generated in the brain and what they reflect [1]. In fact, Gamma was dismissed as 'spare brain noise' at first. There are many speculations about the role of Gamma

waves, some researchers think that they modulate perception and consciousness and that they relate to expanded consciousness and spiritual emergence [6], but nothing has been confirmed yet.

1.4 Brain disorders

The brain is the control center of the body. It is the major part in the nervous system, which controls everything from the human senses to the muscle movements of the body. Therefore, when the brain is damaged, it can lead to severe consequences and different disorders that vary according to which area is affected. Some of these disorders are Dementia, Alzheimer, Epilepsy and schizophrenia.

1.4.1 Dementia

Dementia is an overall term that covers a wide range of specific medical conditions characterized by a decline in memory, language, problem-solving and other thinking skills that affect a person's ability to perform everyday activities [7]. Dementia is caused by damage to or loss of nerve cells and their connections in the brain. It affects people differently depending on the damaged area [8].

1.4.2 Alzheimer

Alzheimer is a type of dementia that causes problems with memory, thinking and behavior. It is a progressive condition such that it develops gradually and becomes more severe within the years. Alzheimer leads to nerve cell death and tissue loss throughout the brain. Over time, the brain shrinks dramatically, affecting nearly all its functions [9].

1.4.3 Epilepsy

Epilepsy is a chronic disorder in which clusters of nerve cells, or neurons, in the brain sometimes signal abnormally causing strange sensations, emotions, and behavior, or convulsions, muscle spasms, and loss of consciousness in some cases [10]. It is characterized by unpredictable seizures. They may be related to a brain injury or a family tendency, but often the cause is completely unknown.

1.4.4 Schizophrenia

Schizophrenia is a chronic and severe mental disorder. People with schizophrenia may seem like they have lost touch with reality. Scientists have found that people who have the disorder may be more likely to have glitches in their genes that may disrupt brain development.

Doctors also believe the brain loses tissue over time. Some imaging tools have shown that people with schizophrenia have less grey matter [11].

1.5 Electroencephalogram

Electroencephalogram (EEG) is an electrophysiological technique used to evaluate the brain function by measuring the electrical activity produced by the ionic current within the neurons of the brain. It records patterns of activity during rest and in response to certain stimuli. The electrical data is recorded from sensors (electrodes) placed at the scalp surface.

1.5.1 Early History of EEG

The electrical properties of the brain were first discovered by the English scientist, Richard Caton (1842-1926). He recorded the electrical activity from the brains of animals using a sensitive galvanometer, noting fluctuations in activity during sleep and absence of activity following death [12]. It was first applied to humans in the 1920s by German neurologist Hans Berger (1873–1941) [1]. He was the one to invent the electroencephalogram and give it its current name [13]. Berger's findings were independently confirmed in early 1934 by Lord Adrian in England and by Hallowell Davis at Harvard, in the United States [14]. The field of clinical EEG began in 1935, after Gibbs, Davis and Lennox described interictal spike waves and the three cycles/s pattern of clinical absence seizures. In 1947, The American EEG Society was founded and the first International EEG congress was held [13].

1.5.2 The 10-20 System

The 10-20 system is one of the two most common systems for defining and naming electrode locations/positions along the scalp. The electrodes are placed at 10% and 20% points along lines of longitude and latitude. In the 10-20 system, electrode names begin with one or two letters indicating the general brain region or lobes where the electrode is placed, and it ends with a number or letter indicating the distance to the midline. Odd numbers are used in the left hemisphere, even numbers in the right hemisphere. Larger numbers indicate greater distances from the midline, while electrodes placed at the midline are labeled with a "z" for zero. Figure 1.3 shows an electrode distribution based on the 10-20 system over the scalp. The Nasion (Nz) indicates the depression between the eyes at the top of the nose. The Inion (Iz) indicates the bump at the back of the head [1].

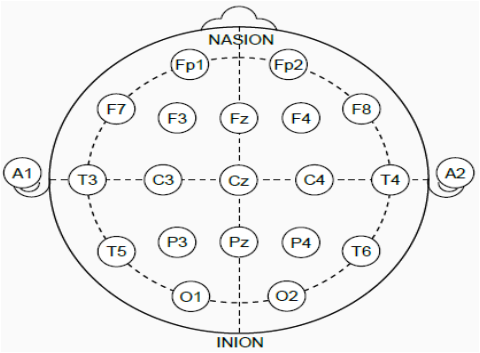


Figure 1.3 Electrodes placement according to the 10-20 system [1].

1.5.3 EEG compared to other imaging techniques

EEG has many advantages over other imaging techniques. The central benefit of EEG is its very high time resolution, which makes it able to capture cognitive and emotional processes and the physiological changes underlying them much better than other brain imaging techniques (such as MRI or PET scanners). Moreover, the process of measuring the neural activity is direct and simple, unlike fMRI in which it is obtained indirectly and requires a much deeper understanding of the relationship between what is measured and how it relates to cognitive processing. EEG is also inexpensive and lightweight, which makes it easily accessible, portable and easy to handle [1]. Despite all the advantages of the electroencephalogram, it has its own limitations, such as its poor spatial resolution which makes it inadequate in many applications. For any distribution of electrical field strengths detected on the scalp, there is an infinity of possible spatial patterns, of electrical activity, that could have generated it [15]. Table 1.1 illustrates the main differences between EEG, MRI, and fMRI [16].

Table 1.1 Comparison between EEG, MRI and fMRI.

	EEG	**MRI**	**FMRI**
Temporal resolution	High	Low	Low
Spatial resolution	Low	High	High
Brain activity measurement	Directly	Only the structure	Indirectly
Level of expertise needed	Some training	Extensive training	Extensive training
Cost	Accessible to many researchers	Requires extensive funding	Requires extensive funding
Portability	Both fully portable and semi portable devices available	Not portable	Not portable

1.6 EEG for Epilepsy diagnosis

EEG is the most common test used to diagnose Epilepsy. The electrodes attached to the scalp, with a paste-like substance or cap, record the electrical activity of the brain. If a person has Epilepsy, it is common to have changes in the normal pattern of brain waves, even when there is no seizure. However, the changes are more noticeable during seizure activity. The doctor may monitor patients on video when conducting an EEG while they are awake or asleep, to record any seizures they experience in order to determine their kind. The test may be done in a doctor's office or the hospital. If appropriate, an ambulatory EEG, which the patient wears at home, may be used. The EEG records seizure activity over the course of a few days. The doctor may give some instructions to trigger the seizures [17]. Recently, many researches are conducted in order to make the process of detecting Epilepsy automatic by means of machine learning. That is also the topic of interest in our project.

1.7 Summary

In this chapter, we briefly described the brain anatomy, its main rhythms and some diseases that result from occasional damages that could occur in it, from which Epilepsy is the most common and the disorder of interest in our project. The EEG technique is presented next with a brief history, the most common system used for the electrodes placement and a comparison with other imaging techniques. Finally, we described the traditional way to diagnose Epilepsy with EEG.

Chapter 2

EEG Signal Feature Extraction

2.1 Introduction – Literature review

Electroencephalography (EEG) records brain activities by measuring the voltage fluctuation on the scalp. This signal has a great potential for diagnosis and treatment of brain disorders. However, it is very difficult to get useful information from raw EEG signals directly. Hence, preprocessing and feature extraction steps are necessary in the EEG signal analysis.

Numerous methods of feature extraction and classification have been proposed throughout the years. The Bonn database is used as a benchmark data set in many of the cited works. It consists of five sets denoted A, B, C, D and E. Sets A and B recordings belong to healthy subjects. Sets C and D recordings belong to epileptic patients during seizure-free intervals. Set E corresponds to seizure recordings (more details about the dataset are provided in chapter 4). Gandhi et al. (2011) [18] used the DWT to extract three features from the EEG signals, energy, entropy and standard deviation. As classifiers, they used SVM and Probabilistic Neural Networks (PNN) to obtain a maximum accuracy of 95.44% for the ABCD-E case [19]. Nicolaou et al. (2012) [20] extracted a single feature, which is the permutation entropy from EEG signals and used the SVM classifier to report 93.5% accuracy for the A-E data sample whereas the maximum accuracy for other data samples such as B-E, C-E, D-E and ABCD-E is 86.1% [19]. M. Z. Parvez and M. Paul (2013) [21] presented an approach based on the high-frequency components of The DCT for feature extraction, which are combined with the bandwidth feature extracted from the Empirical Mode Decomposition (EMD). They used the Least Square SVM (LS-SVM) classifier to identify the ictal and interictal period of epileptic EEG signals from different brain locations. The maximum achieved accuracy on the Freiburg database was 79%. V. Bajaj and R. B. Pachori (2013) [22] proposed a novel method to detect the seizures using the Hilbert transformation of Intrensic Mode Functions (IMFs). The classification achieved an accuracy of 90% [23]. R. J. Martis et al. (2013) [24] used a decision tree classifier with energy, fractal dimension and entropy as features. The achieved accuracy is 95.7% [23]. N. Ahammed et al. (2014) [25] used the Daubechies order 2 wavelet to extract the coefficients. The parameters fed to a linear classifier are energy, entropy, mean, maximum and minimum. They used three sets from the Bonn database, set A, set D and set E. The overall accuracy obtained is 84.2%. Juarez-Guerra et al. (2015) [26] extracted statistical features such as mean, median and variance from the EEG signals and used the feed-forward neural networks to report an accuracy of 93.23% [27]. Zakariya Lasfer et al. (2017) [27] used only sets A and E from the Bonn database for seizure detection. They extracted the wavelet coefficients as features and calculated the energy of each wavelet coefficient. They obtained a maximum

accuracy of 98.1%, a sensitivity of 97.8% and a specificity of 98.1% with the ANN classifier. A. B. Peachap and D. Tchiotsop (2019) [28] decomposed the EEG signal using Laguerre polynomials based wavelets. They reduced the dimensionality with Principal Component Analysis (PCA) and performed the classification using SVM and pattern recognition ANN. They tested multiple cases from the Bonn database. The lowest classification accuracy obtained with ANN was 94% and with SVM, it was 90%, which corresponds to data sample C-E. The best classification accuracy with ANN was 100% and with SVM, it was 98%, which corresponds to data sample B-E. They also pointed out that the scheme they used constitutes a classical case of overfitting, such as all the reported accuracies were 100% before the cross-validation.

2.2 Methodology

In our study, we will use for the preprocessing step a Butterworth low-pass filter to correct and remove artifacts. For feature extraction step, three methods are proposed in this chapter. The first one is to directly extract eight features from the original signal. In the second method, features are extracted from the EEG signal after transforming it to frequency domain using Discrete Cosine Transform. For the third method, the EEG signal is transformed to time frequency domain using Discrete Wavelet Transform then some features are extracted from it. These two steps of EEG signal analysis are described in this chapter.

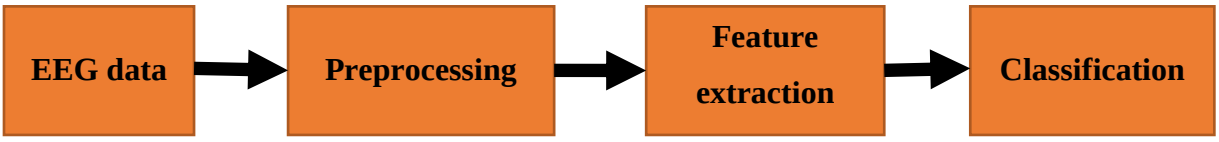


Figure 2.1 Block diagram of the basic steps applied to EEG signal analysis.

2.3 Preprocessing

EEG recording is highly susceptible to various forms of noise and artifacts, such as blinking or muscle movement, that can contaminate the data and distort the picture. So, an initial task of any EEG data analysis is noise and artifact removal, which consists of separating the relevant neural signals from random neural activity that occurs during EEG recordings. This is done in the step of preprocessing, which is a procedure of transforming data into a format that is more suitable for further analysis and interpretable for the user [29]. For this preprocessing step, a filtering is done using a second-order low-pass Butterworth filter to cut off all the frequencies above 60Hz which are viewed as noise.

2.3.1 Butterworth filter

The Butterworth filter, described by the British engineer and physicist Stephen Butterworth, is a type of signal processing filter designed to have frequency response as flat as possible in the passband [30]. It is specified by two parameters, the cutoff frequency and the filter order. When the frequency increases, the frequency response decreases monotonically, and as the filter order increases, the transition band becomes narrower [31] as it is illustrated in figure 2.2.

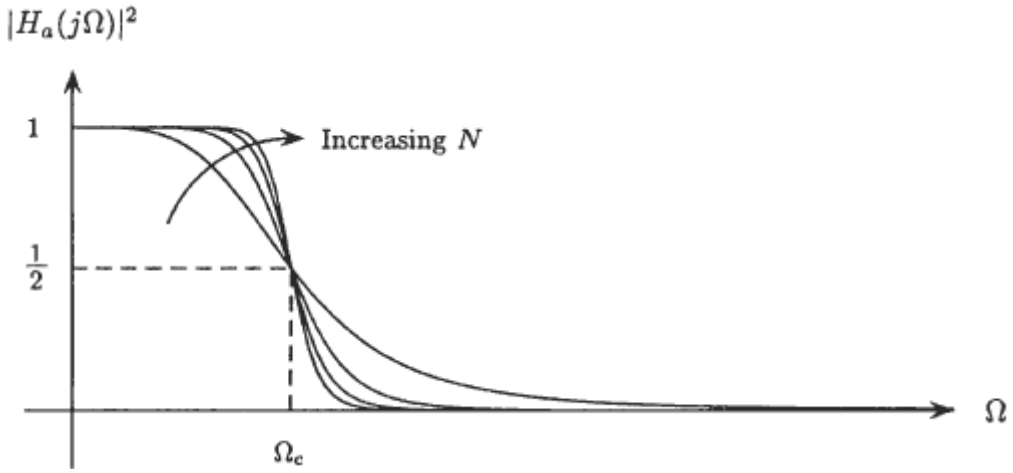
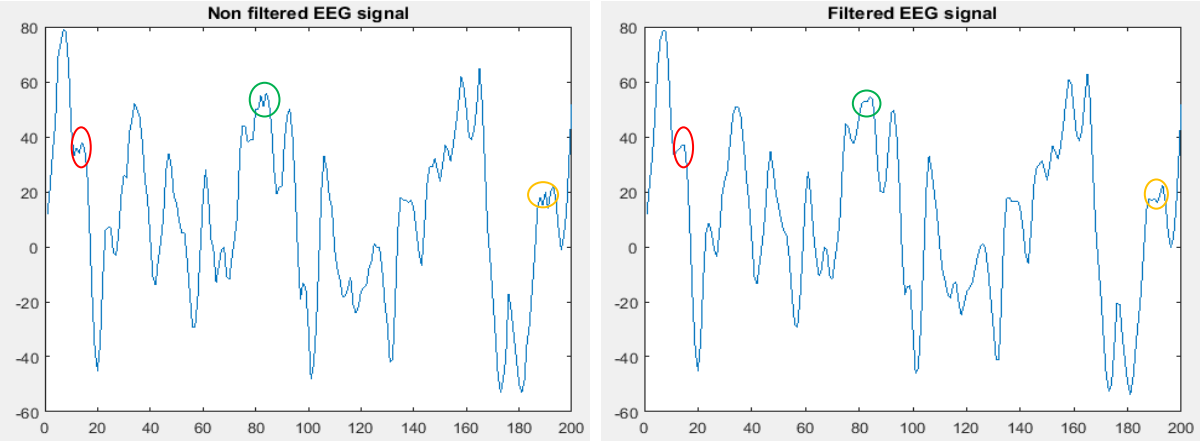


Figure 2.2 The squared magnitude of the frequency response for Butterworth filters of orders N = 2,4,8 [31].

The magnitude-squared function of an N^{th} order Butterworth low-pass filter and a cutoff frequency Ω_c is given by [31]:

$$| \mathrm{H_a}\,(j\Omega)\,|^2 = \frac{1}{1+(\frac{j\Omega}{j\Omega_c})^{2N}} \qquad (2.1)$$

Figure 2.3 shows two recordings of a 200 points EEG signal, where the x-axis represents the sample number and the y-axis represents the amplitude in microvolts. The recording in Figure 2.3(a) is not filtered; while the one in Figure 2.3(b) is filtered using a Butterworth filter of order 2 at a cutoff frequency of 60 Hz. Significant differences can be observed between the two recordings.



(a) (b)

Figure 2.3 One portion of an EEG signal (a) before and (b) after filtering process.

2.3 Feature extraction

After the preprocessing stage, a filtered EEG signal suitable for extracting the needed features is obtained. In this study three methods of feature extraction are used. In the first method, we extract statistical features directly from the filtered time-domain signal. In the second method, we transform the signal to the frequency domain using DCT. While in the third method, the signal is transformed to the time frequency domain by the DWT. This section describes the theoretical background behind these methods.

2.3.1 Feature extraction using statistical parameters

Throughout our study, eight statistical features have been introduced. They are maximum, mean, standard deviation, median, mode, first quartile, third quartile and interquartile range.

- **Mean**

The mean, or average, of a set of values is the ratio of their sum divided by the total number of values in the set. Thus, if there are a total of N numbers in a data set whose values are given by a group of x-values, then the arithmetic mean of these values, denoted by μ, can be found using this formula:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad (2.2)$$

Where x_n is the value of the n^{th} point of the dataset.

- **Standard deviation**

The standard deviation is a statistic value that measures the dispersion of a dataset relative to its mean. It is calculated by determining the variation between each data point relative to the mean. If the data points are further from the mean, there is a higher deviation within the data set. Thus, the more spread out the data, the higher the standard deviation [32]. The standard deviation, σ, is the square root of the variance, σ^2. Unlike the variance which is expressed in squared units, the standard deviation is expressed with the same units as the original data. The standard deviation of a dataset of N values is calculated by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} \tag{2.3}$$

Where x_i is the value of the i^{th} point of the dataset and μ is the mean value of the dataset.

- **Median**

The median of a set of data is the middle number of the set when these numbers are sorted in an ascending or descending order. If there is an odd number of data, the middle value is picked as the median. But if the number of data is even, then there is no single middle value; the median is then defined to be the mean of the two middle values [33]. The median of a dataset of N elements, x, is given by:

$$M = \begin{cases} x_{\frac{N+1}{2}} & \text{if } N \text{ is odd} \\ \frac{1}{2}\left[x_{\frac{N}{2}} + x_{\frac{N}{2}+1} \right] & \text{if } N \text{ is even} \end{cases} \tag{2.4}$$

- **Mode**

The mode is a statistical measure of central tendency in a dataset. It is defined to be the number that occurs most frequently in the set. Thus, the mode is not unique for a particular dataset; there may be several modes in a dataset or none at all [34].

- **First quartile and third quartile**

The first and third quartiles are descriptive statistics that are measurements of position in a data set. For a sorted dataset in an ascending order, the lower half of this dataset is the set of all values that are to the left of the median, and the higher half is the set of all values that are to the right of the median. The first quartile, denoted by Q_1, is the median of the lower half of the data set and the third quartile, denoted by Q_3, is the median of the upper half of the data set [35].

- **Interquartile range**

The interquartile range, IQR, also called the midspread, is a measure of statistical dispersion of a dataset. It is equal to the difference between the third quartile and the first quartile [36].

$$IQR = Q3-Q1 \tag{2.5}$$

2.3.2 Feature extraction using Discrete Cosine Transform (DCT)

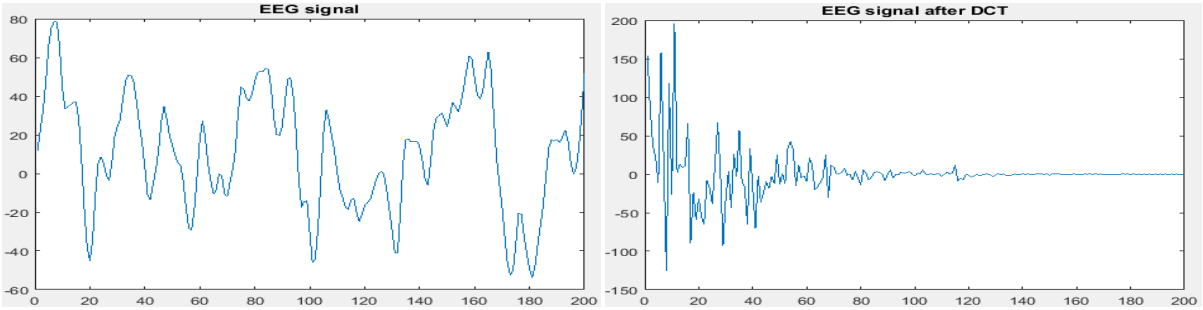
The Discrete Cosine Transform (DCT) is very similar to the Fourier Transform (FT), but DCT involves the use of just Cosine functions and real coefficients, whereas FT makes use of both Sine and Cosine functions and requires the use of complex numbers. Both FT and DCT are transformation methods used for converting a time series signal into basic frequency components and their respective inverse functions convert things back the other way. A DCT expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies. An important feature of DCT is that it takes correlated input data and concentrates its energy in just the first few transform coefficients. If the input data consists of correlated quantities, then only the first few coefficients are large and the other coefficients are zeros or small numbers. Therefore, they can be negligible. The one-dimensional DCT for a signal is given by [37]:

$$G_f = \sqrt{\frac{2}{n}}\, C_f \sum_{t=0}^{n-1} p_t \cos\left(\frac{(2t+1)f\pi}{2n}\right) \tag{2.6}$$

Where $C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0, \\ 1, & f > 0, \end{cases}$ for $f = 0, 1, \ldots, n-1$.

The input is a set of n data values p_t, and the output is a set of n DCT transform coefficients (or weights) G_f.

Figure 2.4(a) shows a 200 points EEG signal in time domain. While figure 2.4(b) shows the same signal after applying DCT on it. In frequency domain, figure 2.4(b), we can see that the energy is compressed into the first coefficients.



(a) (b)

Figure 2.4 An EEG signal (a) before and (b) after DCT.

2.3.3 Feature extraction using Discrete Wavelet Transform (DWT)

- **Overview of wavelets**

In signal processing, there are several ways to extract desirable features from a signal whether on time-domain, frequency-domain or time-frequency domain. Fourier Transform is probably the most popular transform used to obtain the frequency spectrum of a signal. The Fourier coefficients of the transformed function represent the contribution of each sine and cosine function at each frequency [38]. Because time resolution is totally lost, FT is only suitable for stationary signals, i.e., signals whose frequency content does not change with time. However, most of the biological signals, such as, Electrocardiogram, Electromyography, etc have different characteristics at different time or space, i.e., they are non-stationary. Therefore, in the analysis of these signals, both frequency and time information are needed simultaneously [39]. To solve this problem, the Short-Time Fourier Transform (STFT) was introduced. In STFT the non-stationary signal is divided into small portions, which are assumed to be stationary. This is done using a window function of a chosen width, which is shifted and multiplied with the signal to obtain the small stationary signals. The Fourier Transform is then applied to each of these portions to obtain the STFT of the signal [39]. By doing this, we get a two-variable signal as a function of time and frequency. The drawback is that once the particular size for the time window is chosen, that window is the same for all frequencies. The Wavelet Transform (WT) came to solve the problem of FT and STFT. In contrast to STFT, which uses a single analysis window, the Wavelet Transform uses long windows where low-frequency information is needed, and shorter windows where high-frequency information is needed [40]. The Wavelet Transform provides accurate frequency information at the low frequencies and accurate time information at high frequencies. This property is important in biomedical applications, because most signals in the biomedical field always contain high frequency components with short time periods and low frequency components with long time periods.

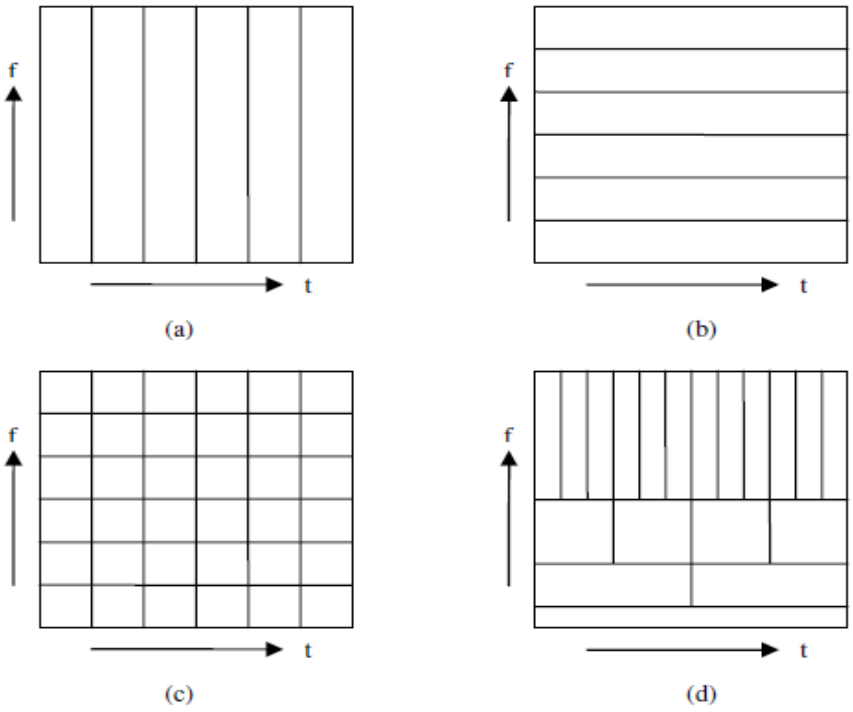


Figure 2.5 The Time-Frequency tiling for (a) Time-Domain (b) Frequency-Domain (c) STFT (d) DWT [39].

In figure 2.5(a) no frequency information is given, only the time-frequency tiling in the time-domain plane is shown. Figure 2.5(b) shows the tiling in frequency-domain plane, it does not give any time information. Similarly figure 2.5(c) shows the tiling in STFT and figure 2.5(d) shows the tiling in Wavelet Transform. Figure 2.5(c) and (d) show that STFT gives a fixed resolution at all times, whereas Wavelet Transform gives a variable resolution [39].

- **Continuous Wavelet Transform**

The Continuous Wavelet Transform (CWT) is a linear operation that decomposes a continuous signal, using a continuous wavelet function, into wavelet coefficients that are functions of scale and position [40]. To classify an analyzing function $\Psi(t)$ as a wavelet, it must satisfy the following criteria [41]:

a) A wavelet must have finite energy:

$$E = \int_{-\infty}^{+\infty} |\Psi(t)^2| \, dt < \infty \tag{2.7}$$

b) If $\Psi(f)$ is the Fourier transform of the wavelet $\Psi(t)$, the following condition must hold:

$$C_{\Psi} = \int_{0}^{+\infty} \frac{|\Psi(f)|^2}{f} \, df < \infty \tag{2.8}$$

This condition implies that the wavelet has no zero frequency component ($\Psi(0) = 0$), i.e. the mean of the wavelet $\Psi(t)$ must be equal to zero.

c) For complex wavelets, the Fourier transform $\Psi(f)$ must be both real and vanish for negative frequencies.

The CWT of a signal $f(t)\ \epsilon\ L^2(R)$ at scale a and translation τ is given by [42]:

$$W_f\ (a, \tau) = \omega(a)\ \int_{-\infty}^{+\infty} f(t)\ \Psi^*\left(\frac{t-\tau}{a}\right)dt \qquad (2.9)$$

The symbol * denotes the complex conjugation. $\omega(a)$ is a weighting function used for energy conservation purpose. $\omega(a)$ is set to $\frac{1}{\sqrt{|a|}}$ to ensure that the wavelet would have the same energy on all the scales. Equation (2.8) shows that the signal to be analyzed, $f(t)$, is correlated with stretched/dilated copies of the mother wavelet $\Psi(t)$. For small values of a, the wavelet is contracted and the transform gives information about the finer details of $f(t)$. For large values of a, the wavelet expands and the transform gives a coarse view of the signal [42]. However, the major weakness of CWT is that scaling parameter a and translation parameter τ change continuously. Thus, the coefficients of the wavelet for all available scales after calculation will consume a lot of effort and yield a lot of unused information [43].

- **Discrete Wavelet Transform**

In order to address the weakness of the CWT, discrete wavelet transform (DWT) has been defined. DWT is a kind of wavelets that restrict the value of scale and translation. The restriction is like the scale is increasing in the power of 2 (a = 1, 2, 4, 8,…) and the translation is the integer (τ = 1, 2, 3, 4, …). An efficient way to implement this is by passing the signal through a Low-Pass Filter (LPF) and a High-Pass Filter (HPF), dividing it into a lower and upper frequency bands. Then, the lower band is subsequently divided into a second level, lower and upper bands. The process is repeated, taking the form of a binary or "dyadic" tree. The lower band is referred to as the "Approximation" and the upper band as the "Detail"[44].

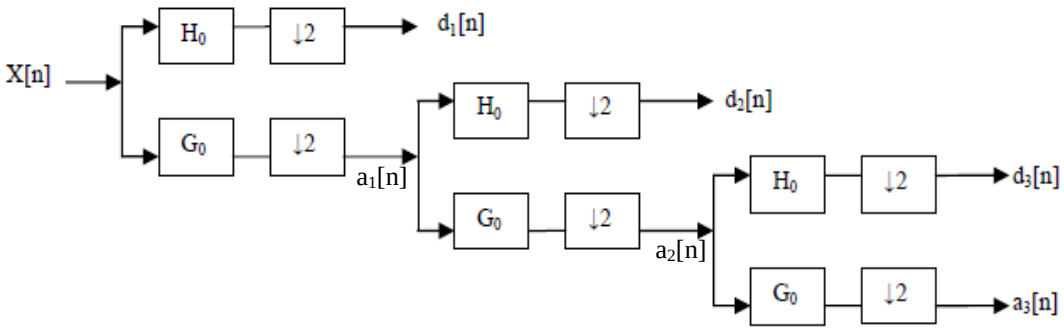


Figure 2.6 Three-level wavelet decomposition tree [39].

Figure 2.6 represents the procedure of the DWT decomposition for an input discrete time-domain signal x[n]. The DWT is computed by successively passing x[n] through a series of

low-pass and high-pass filters. Each stage consists of two digital filters and two down-samplers by 2 to generate the digitized signal. The first filter, H_0, is the discrete mother wavelet, which is a high-pass filter, and the second, G_0, is a low-pass filter. The downsampled outputs of the first high-pass filter produce the detail information $d_1[n]$, while the downsampled outputs of the first low-pass filter produce the coarse approximation, $a_1[n]$. The first approximation, $a_1[n]$, is again decomposed and this process is repeated at each stage [39]. The decomposition of the signal can be mathematically expressed as follows [45]:

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k].H_0[2n-k] \quad (2.10)$$

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k].G_0[2n-k] \quad (2.11)$$

- **Wavelet families**

  In this section, we will present the six wavelet families used in our experiments.

  a. Haar wavelet

Haar wavelet was the first mother wavelet proposed by Alfred Haar, it has the shortest length of support among all orthogonal wavelets. This wavelet is conceptually simple and fast, however the technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions. Haar wavelet represents the Daubechies 1 wavelet function [46].

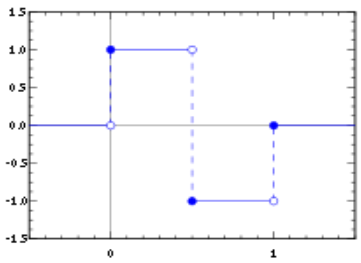


Figure 2.7 Haar wavelet (x axis is time and y axis is amplitude) [46].

b. Daubechies wavelets

Daubechies family of wavelets, abbreviated as db, is based on the work of Ingrid Daubechies. It is a family of orthogonal wavelets defining a DWT and characterized by a maximum number of vanishing moments for some predefined support length. With each wavelet, there is a corresponding scaling function generating an orthogonal multiresolution analysis [47]. The name of the Daubechies wavelet is represented as dbN where N gives its order. The order N refers to the number of vanishing moments which is equal to half the length of the support [48]. Wavelet functions of db2 to db9 are shown in figure 2.8.

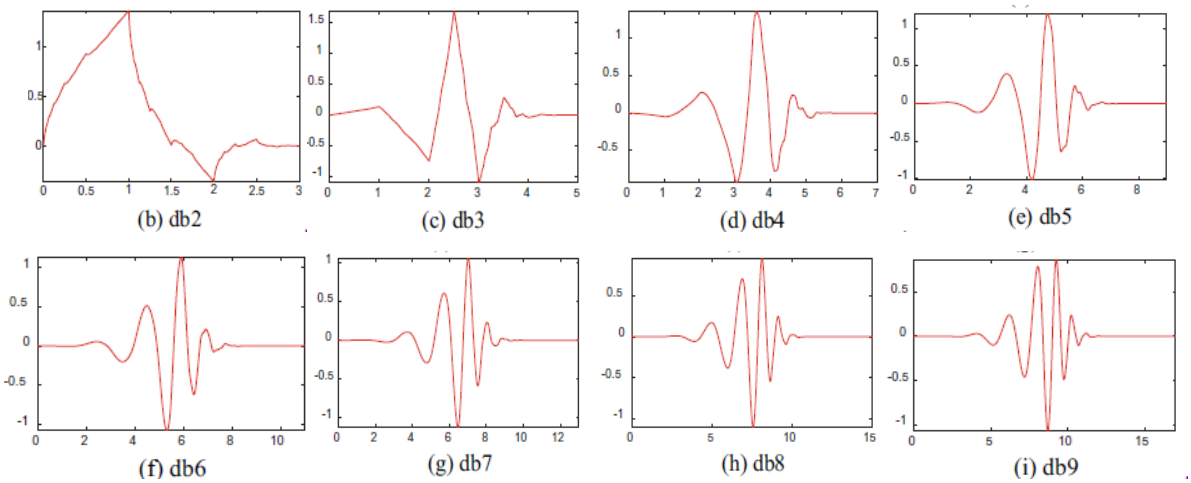


Figure 2.8 Wavelet functions of Daubechies family (x axis is time and y axis is amplitude) [47].

c. Biorthogonal wavelets

A biorthogonal wavelet, abbreviated as bior, is a wavelet where the associated wavelet transform is invertible but not necessarily orthogonal. The advantage of biorthogonal wavelet transform is that it allows more degrees of freedom compared to the orthogonal wavelet [49]. A biorthogonal wavelet has two scaling functions and two wavelet functions. One scaling/wavelet function is for decomposition, the other is for reconstruction [47]. All the functions of bior2.2 wavelet are given in figure 2.9.

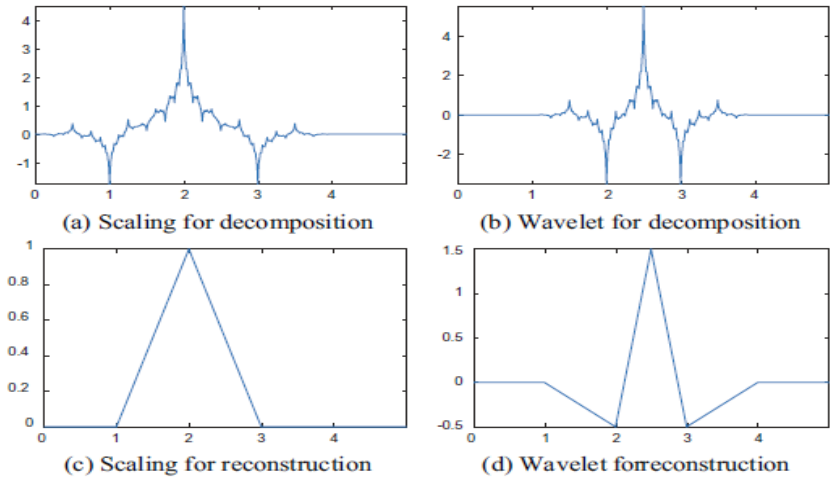


Figure 2.9 Functions of the bior2.2 wavelet (x axis is time and y axis is amplitude) [47].

d. Coiflet wavelets

Coiflets, abbreviated as coif, are discrete wavelets designed by Ingrid Daubechies and Ronald Coifman, to have scaling functions with vanishing moments. Their scaling function has N/3−1 vanishing moments, and the wavelet function has N/3 vanishing moments [47].

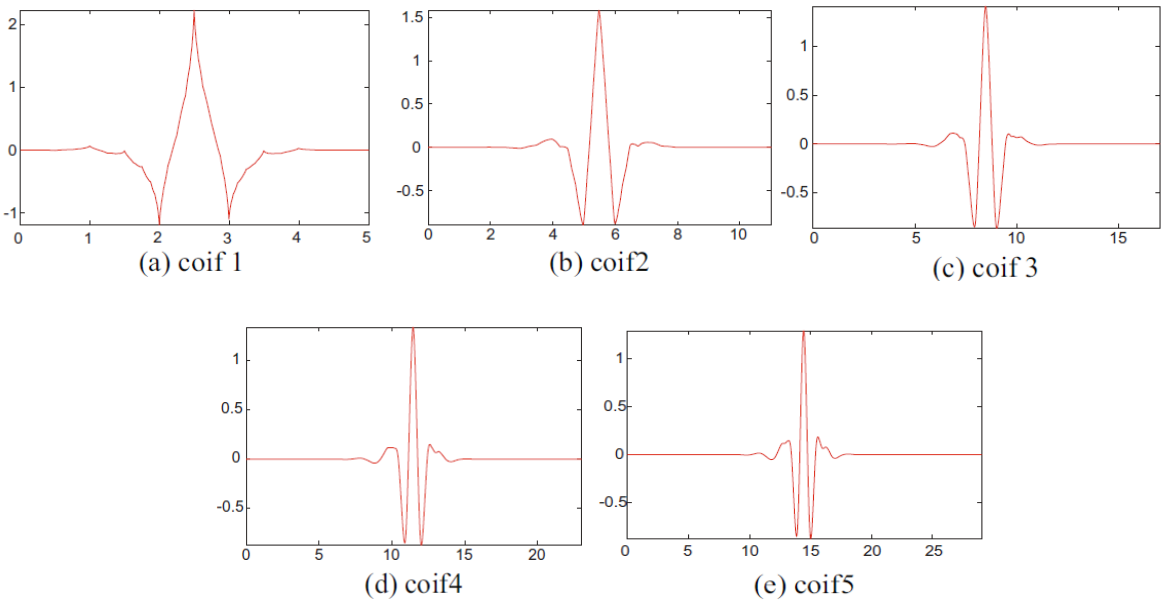


Figure 2.10 Wavelet functions of the Coiflet family (x axis is time and y axis is amplitude) [47].

e. Symlet wavelets

Symlets are also orthogonal and compactly supported wavelets. They are a modified version of Daubechies wavelets with increased symmetry. The associated scaling filters are near linear-phase filters. The properties of Symlets are nearly the same as those of the db wavelets [50]. The Symlet wavelets for orders 2 up to 8 are shown in figure 2.11.

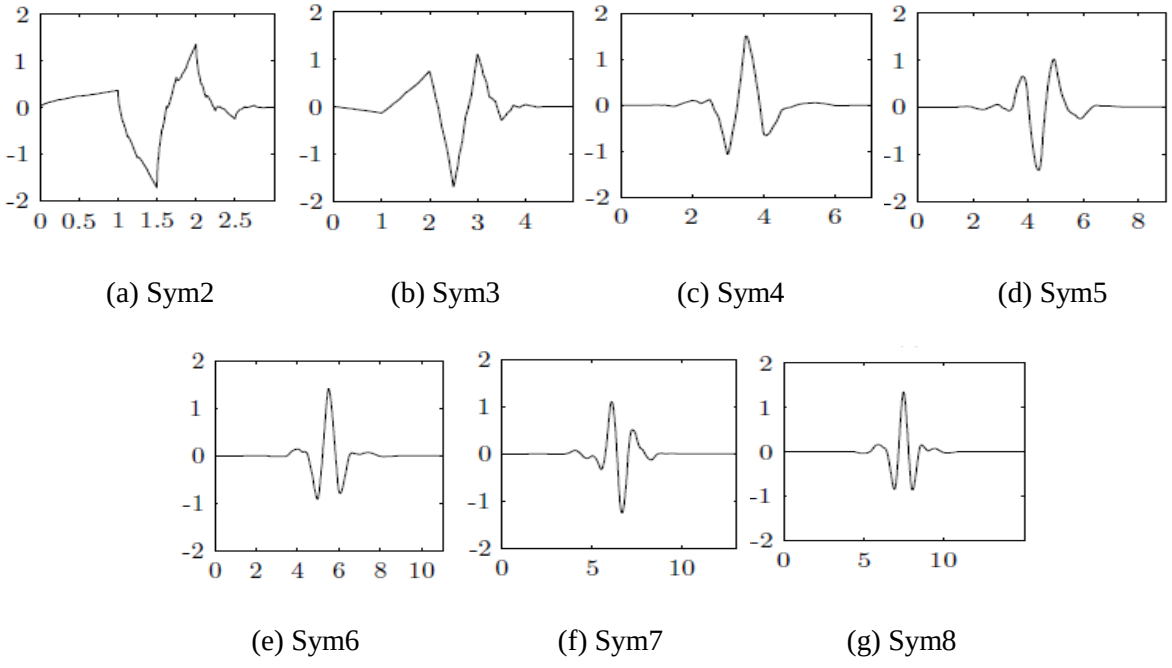


Figure 2.11 Wavelet functions of the Symlet family (x axis is time and y axis is amplitude) [50].

f. Discrete Meyer wavelet

Meyer wavelets are orthogonal wavelets proposed by Yves Meyer. Their scaling wavelet function is symmetric. Meyer is band limited and has infinite number of supports [48]. Discrete Meyer is a discrete approximation of the Meyer wavelet.

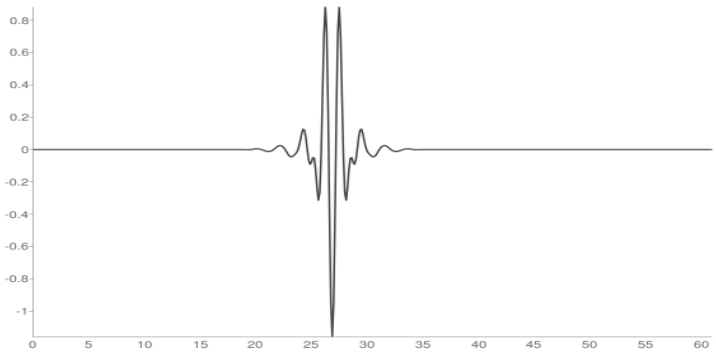


Figure 2.12 Discrete Meyer wavelet (x axis is time and y axis is amplitude).

2.4 Summary

This chapter briefly described two steps of EEG signal analysis. First, a preprocessing of EEG signal using a low-pass filter has been introduced, then three feature extraction methods have been deeply presented.

Chapter 3

EEG Signal Classification

3.1 Introduction

Signal classification means to analyze different characteristic features of a signal, and based on them, decide to which grouping or class the signal belongs. The resulting classification decision can be then mapped back into the physical world to reveal information about the physical process that created this signal. In order to have a broad understanding of classification, this chapter mainly provides an overview of machine learning and classification algorithms.

3.2 Machine Learning

Machine Learning is a branch of artificial intelligence based on the idea that systems can automatically learn and improve from experience without being explicitly programmed. The process of learning begins with observations or training data in order to look for patterns in that data and make better decisions in the future based on the provided data [51]. There are three types of learning approaches, namely, supervised, unsupervised and reinforcement learning. In a nutshell, reinforcement learning is dynamic programming that trains algorithms using a system of reward and punishment. Unsupervised learning is when the model is given training based on unlabeled data without any guidance while in supervised learning, the machine learns from a labeled dataset with guidance. This latter is deeply explained in the following section.

- **Supervised machine learning**

In supervised learning a training dataset which consists of inputs paired with their correct outputs is used to develop a prediction model. The model can then make predictions of the output values for a new dataset. The basic idea of supervised learning is to estimate a function f: $x \rightarrow y$ that maps inputs x to target values y, given a set of N training examples $d=\{(x_i,y_i)\}_{i=1}^{N}$. The goal is to approximate the mapping function so well that output variable y can be predicted for a new input data x. The supervised approach is further divided into:

- **Regression:** It is used to predict continuous values. The task of the Regression algorithm is to find the mapping function that maps the input variable x to the continuous output variable y [52].
- **Classification:** It is the process of finding a model (function) which separates the data into multiple categorical classes based on different parameters. Classification problems consist of taking input vectors and deciding which of N predefined classes they belong to, based on training from exemplars of each class. The most important point about the classification problem is that it is discrete, each example belongs to precisely one class, and the set of classes covers the whole possible output space [53]. Classification can be

binary, where instances are classified into two groups, or multi-class where instances are classified into one of three or more classes.

3.3 Some supervised learning algorithms

Supervised learning uses classification algorithms and regression techniques to develop predictive models. Several algorithms have been developed. In this section, the three algorithms used in the context of our study to perform binary classification are deeply explained.

3.3.1 k-Nearest Neighbor (k-NN)

A. k-NN Theory

The k-nearest neighbor's algorithm is a non-parametric and supervised machine learning method used for classification and regression. In classification, k-NN is based on similarity measure among the training and test sets. Given a point x_0 to be classified into one of N groups, the k nearest data points to x_0 must be found. The classification rule is to assign x_0 to the population that has the most observed data points out of the k nearest neighbors. Points for which there is no majority are either classified to one of the majority populations at random, or left unclassified. The illustration of the k-NN algorithm is given in figure 3.1. The green circle is the sample which is to be classified, blue squares and red triangles are two different classes in the training set. If k=3, then the 3 nearest neighbors, illustrated by the black circle with the solid line, will decide to which class the analyzed sample will be assigned. If k=5, then the 5 nearest neighbors will be considered (illustrated by the black circle with the dotted line) [54].

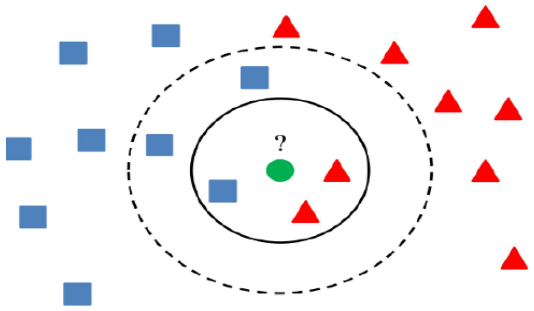


Figure 3.1 Illustration of k-NN algorithm.

The advantage of k-NN classification is its simplicity. There are only two important concepts that should be taken into consideration [55]:

- The parameter k, which decides how many neighbors will be chosen for k-NN algorithm. The appropriate choice of k has significant impact on the diagnostic performance of k-NN algorithm. A large k reduces effect of the noise on the classification, but makes

boundaries between classes less distinct [56]. The key to choose an appropriate k value is to strike a balance between overfitting and underfitting. In binary classification problems, k is generally chosen to be an odd number as this avoids tied votes [56].

- The choice of a method to measure the distance between the attributes in the testing set and the training set. The most significant ones are Euclidean distance and Manhattan distance. The mathematical formula for measuring the Euclidian distance between feature vectors $X=(x_1,x_2,\ldots, x_n)$ and $Y=(y_1, y_2,\ldots, y_n)$ is given below [54]:

$$d(x,y) = d(y,x) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3.1)$$

B. k-NN classification steps

The k-NN classification process is usually based on the following steps [57]:

- Determine parameter k as the number of nearest neighbors.
- Calculate the distance between each testing sample and all the training set element by element.
- Sort the distances and determine the k nearest neighbors.
- Determine the classes of each of the k nearest neighbors.
- Apply majority voting to decide the class of the new data.

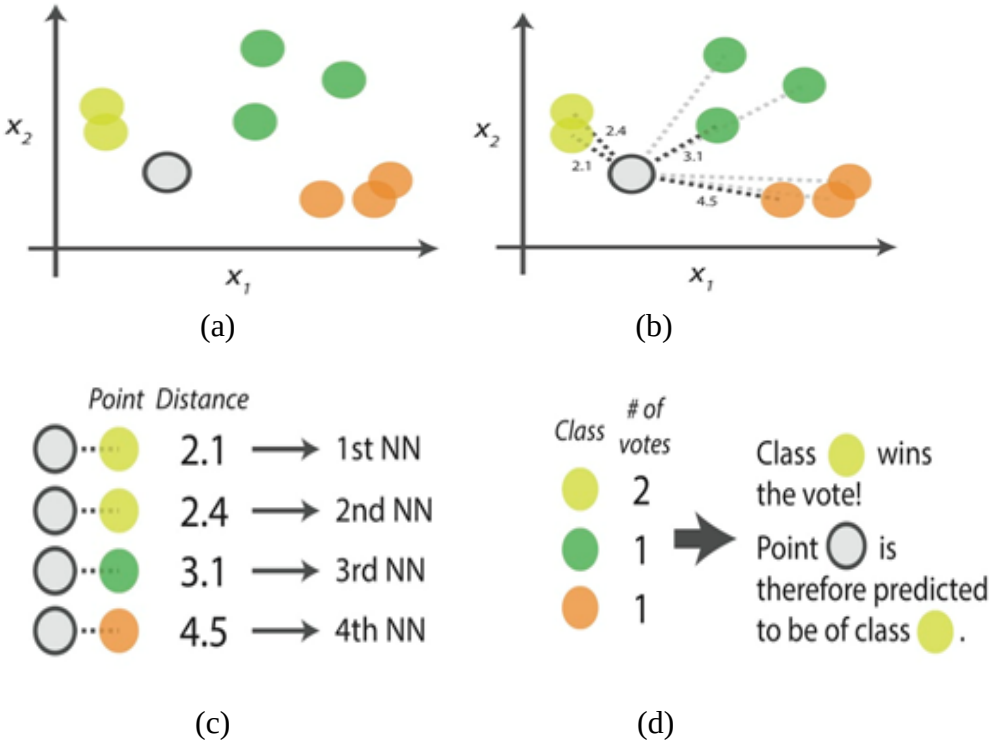


Figure 3.2 Illustration of 4-NN algorithm steps to classify the gray point [58].

3.3.2 Support Vector Machine (SVM)

Support vector machine, or SVM, is a machine learning algorithm initiated by Vladimir Vapnik. It was developed to solve linear or non linear classification and regression problems. The basic idea of the SVM classification algorithms is to construct a hyperplane that separates two groups if possible. The optimal hyperplane must have the largest distance to the nearest training-data points of the two classes in order to reduce the misclassification error. These points are called support vectors and the distance between the hyperplane and the support vectors of each class is called the margin. The goal of the SVM algorithm is to find the optimal separating hyperplane which maximizes the margin [59]. There are two types of SVMs, namely linear SVM and non linear SVM.

A. Linear SVM

Let's assume that we have a training set S which contains N training samples, $x_i \epsilon R^n$ with $i=1,2,...,N$. Each point x_i belongs to either of two classes and thus is given a label $y_i \in \{-1,1\}$. The training set is represented as $\{x_i, y_i\}$. The goal is to establish the equation of an optimal hyperplane that divides S leaving all the points of the same class on the same side [60]. The equation of the optimal hyperplane with a normal vector w and a bias $b \epsilon R$ is given by:

$$w.x + b = 0 \qquad (3.2)$$

The most important task of SVM is to find values of w and b of the hyperplane equation.

- **Hard margin case**

If the training data is linearly separable, two parallel hyperplanes that separate the two classes of data can be selected so that the distance between them is as large as possible.

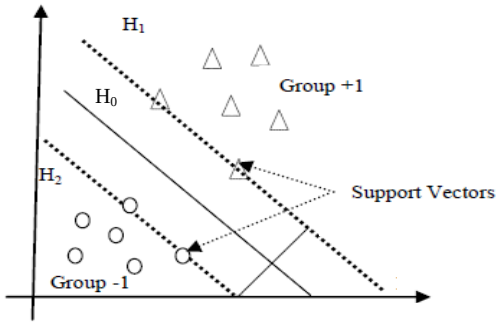


Figure 3.3 Optimal hyperplane in SVMs for linearly separable data [61].

In order for the hyperplane H_0 in figure 3.3 to correctly separate the two groups, the following two constraints which state that each data point must lie on the correct side of the margin must be satisfied [62].

$$x_i \,.w + b \geq +1, \text{ for all } y_i = +1 \tag{3.3}$$

$$x_i \,.w + b \leq -1 \quad , \text{ for all } y_i = -1 \tag{3.4}$$

The margin between the two hyperplanes H_1 and H_2 is geometrically computed to be $\frac{2}{||w||}$ where $||w||$ is the Euclidian norm of the vector w. Therefore maximizing the margin is equivalent to minimizing $\frac{1}{2}||w||^2$, in order to perform Quadratic Programming (QP) optimization later on [62]. After combining equations (3.3) and (3.4) into a single equation, the optimization problem can be formulated in the following way [60]:

$$\begin{cases} \text{Minimize } \frac{1}{2}\,||w||^2 \\ \text{subject to } \; y_i\,(x_i\,.w+b) \geq 1, \qquad \forall\, i \end{cases} \tag{3.5}$$

In order to cater for the constraints in this minimization, Lagrangian multipliers α_i, where $\alpha_i \geq 0\ \forall i$, are introduced, then the primal formulation of this optimization problem becomes [62]:

$$L_P(w, b, \alpha) = \frac{1}{2}\,||w||^2 - \sum_{i=1}^{N} \alpha_i [y_i\,(w.x_i + b) - 1] \tag{3.6}$$

The objective is then to minimize (3.6) with respect to w and b, this simultaneously requires that the derivatives of $L_P(w, b, \alpha)$ with respect to w and b vanish [60].

$$\frac{\partial L_P(w,b,\alpha)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i\, y_i\, x_i \tag{3.7}$$

$$\frac{\partial L_P(w,b,\alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i\, y_i = 0 \tag{3.8}$$

Substituting (3.7) and (3.8) into (3.6) leads to a dual formulation of the optimization problem as a function of $\alpha_i$only, which is to be maximized along with the following constraints [63]:

$$\begin{cases} \text{Maximize } L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i\ \alpha_j\ y_i\ \ y_j\ \ x_i\, x_j \\ \text{Subject to } \sum_{i=1}^{N} \alpha_i\, y_i = 0 \quad \text{and} \quad \alpha_i \geq 0\ \ \forall\, i \end{cases} \tag{3.9}$$

Thus, the dual optimization problem can be solved by QP solver, which will return the coefficients α_i. Each $\alpha_i > 0$ indicates that the corresponding x_i is a support vector. Then, the value of w will be determined from equation (3.7). Note that only instances with $\alpha_i > 0$ are needed to express the value of w [60]. As a final step, with the support vectors, the value of the bias b can be calculated [62]:

$$b = \frac{1}{N_s} \sum_{s \in SV} \left[\, y_s - \sum_{i \in SV} \alpha_i\ \ y_i\, x_i\, x_s \,\right] \tag{3.10}$$

Where SV is the set of the support vectors and N_s is the number of the support vectors.

Now that the optimal separating hyperplane is determined, the decision function is expressed as follows and its sign determines the predicted classification of x [60]:

$$f(x) = w.x + b \tag{3.11}$$

- **Soft margin case**

When the data are not completely separable as in hard margin case, soft margin SVM allows some errors in the classification. SVM, in this case, will classify most of the data correctly, while allowing the model to misclassify a few points in the vicinity of the separating boundary.

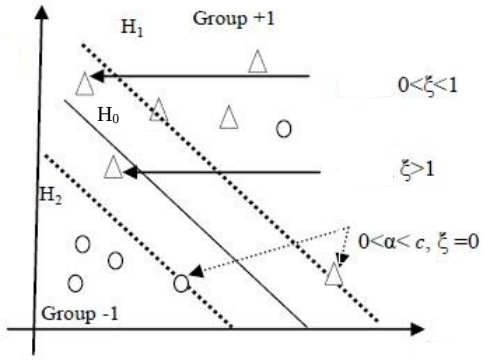


Figure 3.4 Optimal hyperplane in SVMs for linearly non separable data [61].

To handle linearly non separable data and allow for some of the samples to lie inside the margin or even cross further among the instances of the opposite class, positive slack variables ξ_i , $i= 1,2;...,N,$ that relaxes constraints in equations (3.3) and (3.4) are introduced. Based on these new criteria, the relaxed constraints with the slack variables then become [60]:

$$\forall i \begin{cases} x_i .w + b \geq +1 - \xi_i & \text{for } y_i = +1 \\ x_i .w + b \leq -1 + \xi_i & \text{for } y_i = -1 \\ \xi_i \geq 0 \end{cases} \tag{3.12}$$

Now a point x_i can satisfy the constraints even if it is on the wrong side of the decision boundary, as long as ξ_i is large enough. Of course all constraints can be trivially satisfied this way. To prevent this, we penalize the sum of ξ_i [63] and the optimization problem can be reformulated as follow:

$$\begin{cases} \text{Minimize } \frac{1}{2} \|w\|^2 + c \sum_{i=1}^{N} \xi_i \\ \text{Subject to } y_i (x_i .w + b) \geq +1 - \xi_i \\ \text{and } \xi_i \geq 0 \; \forall i \end{cases} \tag{3.13}$$

Where the cost coefficient $c>0$ is a hyperparameter that specifies the misclassification penalty and is tuned by the user based on the classification task and dataset characteristics [60].

The process of optimization is going through the same steps as previous: Lagrangian, optimization of α_i parameters, determining w and b values for classification hyperplane. The dual stays the same, but with additional constraints on α_i parameters: $0 \leq \alpha_i \leq c$ [57].

B. Non linear SVM

When a problem is not linearly separable in input space, soft-margin SVM cannot find an optimal separating hyperplane that minimizes the number of misclassified data points. For that, a mapping function $\Phi(x)$ is used to map the data into a higher-dimensional feature space where data will be linearly separable. By use of a kernel function it is possible to compute the separating hyperplane without explicitly carrying out the mapping into feature space [61]. The equation of the kernel function is given by:

$$K(x, y) = <\Phi(x), \Phi(y)> \tag{3.14}$$

Several kernel functions exist, the two most widely used ones are the polynomial kernel and the Gaussian radial-basis function (RBF) kernel, their equations are given below respectively [60]:

$$K(x, y) = (1 + x.y)^{d} \tag{3.15}$$

$$K(x, y) = exp\left(-\frac{1}{2\sigma^2}\|x - y\|^2\right) \tag{3.16}$$

Modified and enhanced SVM constructs an optimal separating hyperplane in the higher-dimensional space. Hence, the optimization problem becomes [61]:

$$\begin{cases} \text{Maximize } L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{Subject to } \sum_{i=1}^{N} \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq c \ \ \forall i \end{cases} \tag{3.17}$$

Using kernel function, minimization of dual Lagrangian is performed in the feature space. Then, all margin parameters are determined, without representing points in this new space [57]. Finally, the decision function is given by [42]:

$$f(x) = \left[\sum_{i \in SV} \alpha_i \ y_i \ K(x_i, x_j) + b\right] \tag{3.18}$$

Where SV is the set of the support vectors satisfying $0 < \alpha_i < c$.

3.3.3 Artificial Neural Network (ANN)

Artificial neural networks are computing systems, in which a computer learns to perform tasks by analyzing training examples, generally without being programmed with task-specific rules [64]. ANNs take inspiration from the learning process of human brain. This latter is composed of cells called neurons interconnected with links (or axons). Similar to the brain, an ANN is composed of processing units called artificial neurons and interconnections. A graph of a network consists of a number of nodes connected through directional links. Each node

represents a processing unit, and the links between nodes specify the causal relationship between connected nodes [57].

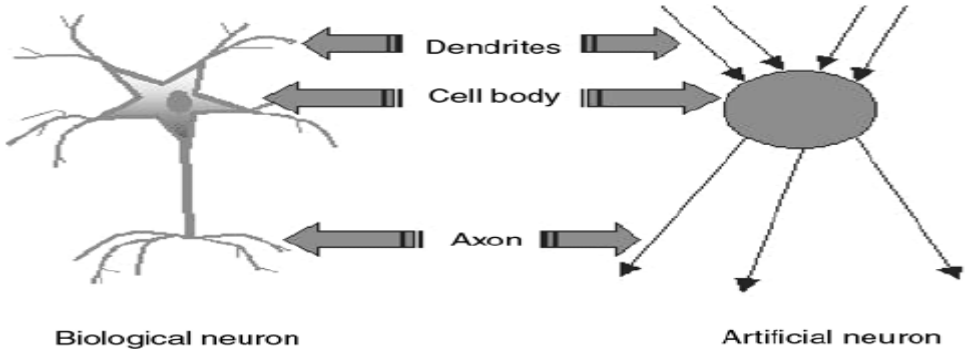


Figure 3.5 Analogy between biological neuron and artificial neuron [65].

A. Model of an artificial neuron

An artificial neuron is an information-processing unit that is the basic computational element of a neural network. It receives input from some other neurons, or perhaps from an external source, processes it, passes it through an activation function and returns the activated output. The block diagram in figure 3.6, which is a model of an artificial neuron, shows that it consists of three basic elements [57]:

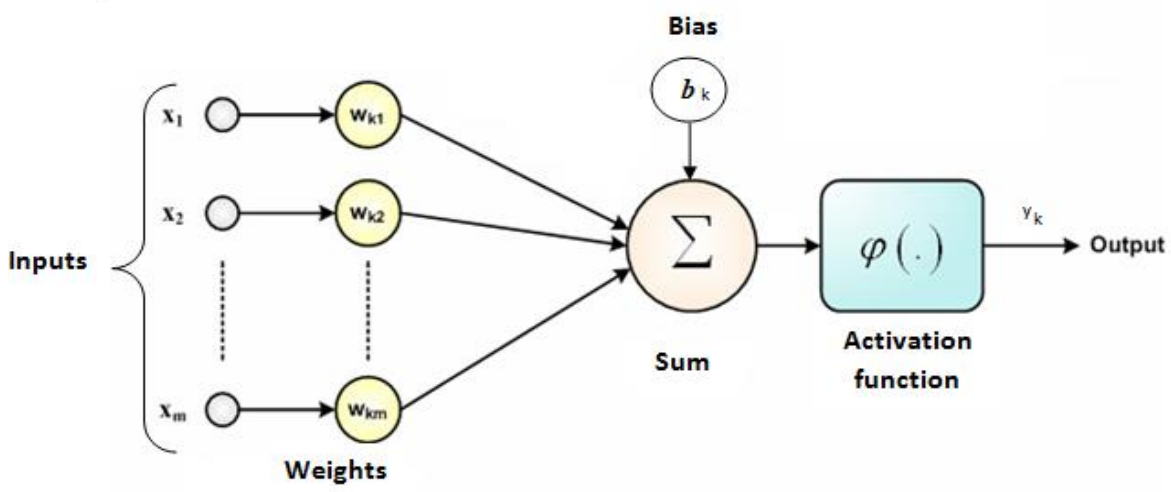


Figure 3.6 Model of an artificial neuron.

- A set of inputs x_i characterized by a weight w_{ki}. Weights may lie in a range that includes positive as well as negative values.
- An adder for summing the weighted inputs and the bias b_k. The bias has the effect of increasing or lowering the input of the activation function, depending on whether it is positive or negative.

- An activation function φ which determines the output y_k according to the following equation

$$y_k = \varphi \left(\sum_{i=1}^{m} x_i \, w_{ki} + b_k \right) \quad (3.19)$$

Activation functions determine whether a neuron should be activated ("fired") or not and help normalize the output of each neuron to a range between 1 and 0 or between -1 and 1. Several forms of activation functions exist; the ones used in our ANN classification method are defined below:

- **Sigmoid:** It is a non-linear activation function, with a smooth gradient, used mostly in binary classification problems, as it takes the input and outputs another value between 0 and 1. Thus, the result can be predicted easily to be 1 if the value is greater than 0.5 and 0 otherwise [66]. This function has two major drawbacks, the first is that it saturates and kills gradients and the second is the fact that its outputs are not zero centered [67]. Sigmoid equation is defined as follows:

$$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (3.20)$$

- **Rectified Linear Unit (ReLU):** It is the most widely used activation function within hidden layers of an ANN. It activates a single node if the input is above a certain threshold. Compared to Sigmoid, this function rectifies vanishing gradient problem and is less computationally expensive [68], as its formula is very simple:

$$\text{ReLU}(x) = \max(0, x) \quad (3.21)$$

B. Architecture of ANNs

A typical neural network has few dozen to hundreds, thousands, or even millions of artificial neurons arranged in a series of layers, each neuron is connected to the layers on either side. ANN normally consists of an input layer, hidden layer(s) and an output layer. An example of ANN structure is shown in figure 3.7. The first layer is the input layer; it receives the inputs provided by the programmer. Number of neurons in this layer corresponds to the number of inputs of the neuronal network. The last layer is the output layer, it contains the same number of neurons as the number of the output values of the network. Between input and output layer, one or several hidden layers may exist. The hidden layers perform all the calculations to find hidden features and patterns.

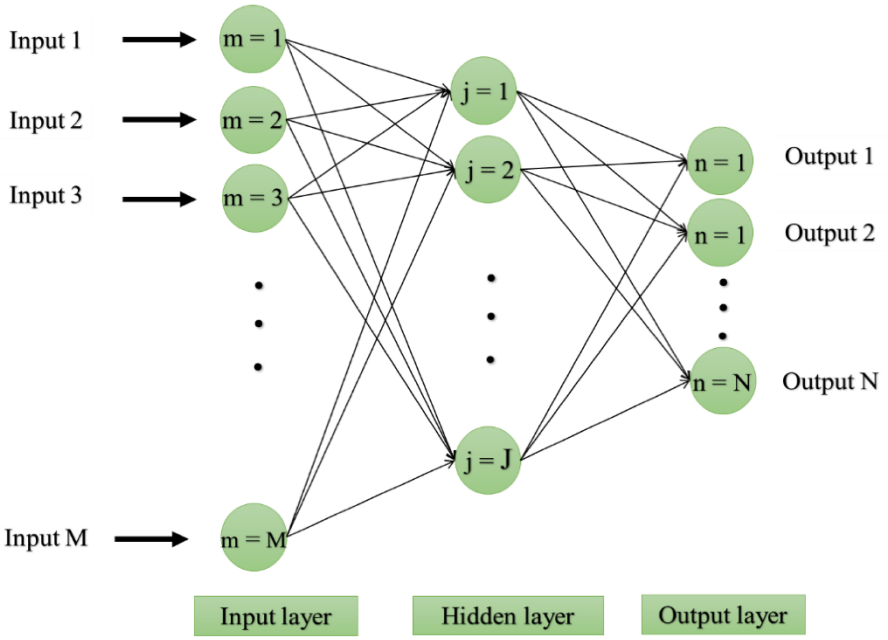


Figure 3.7 ANN architecture (feedforward network).

Neural networks can be divided into two main types:

- Feedforward Networks: They are defined as all networks that do not gain feedback from the network itself. This means that there are no links between nodes in the same layer and the input data flows in one direction, from the input nodes through 0 to n hidden nodes to the output nodes [69]. The most common and basic feedforward networks are Multi Layer Perceptrons (MLPs).

- Recurrent Networks: They are defined as all networks that contain a feedback loop; therefore data from later stages are used for the learning process in earlier stages [69].

C. ANN training process

Among the many interesting properties of a neural network is its ability to learn from its environment and to improve its performance through learning. An ANN learns about its environment by adjusting the weights and biases values of its neurons to minimize the discrepancy between the network output and the desired output. This adjustment process is called training. The most common algorithm in supervised learning is the backpropagation algorithm. Basically, this algorithm consists of two phases performed through the different layers of the network: a forward pass and a backward pass [57]. In the forward pass, weights are initiated to random values and training data are applied to the input nodes of the network. These training data are propagated through the network layer by layer till reaching the output layer where a set of outputs is produced as the actual response of the network. Then, a cost function is used to calculate the error by comparing the actual outputs of the network with the

estimated ones. The cost function used in binary classification is Cross-entropy because its output layer is a probability value between 0 and 1[70]. During the forward phase, the synaptic weights of the network are all fixed. During the backward phase, on the other hand, the weights are all adjusted to make the actual response of the network closer to the desired response [57].

3.4 Summary

This chapter briefly introduced some concepts of machine learning and classification process. Then three supervised machine learning algorithms used in this study have been deeply described.

Chapter 4

Experiments & Results

4.1 Introduction

This chapter describes and compares the performance of three methods, at the level of the feature extraction stage, proposed for Epilepsy detection from EEG signals during both ictal and interictal intervals. The raw EEG signal goes through a preprocessing step, then feature extraction and finally the classification. Figure 4.1 depicts a block diagram, which illustrates the three used methods in more details. The same procedures are used for both experiments. The difference lies in the way we divide the data. All the details are provided later on.

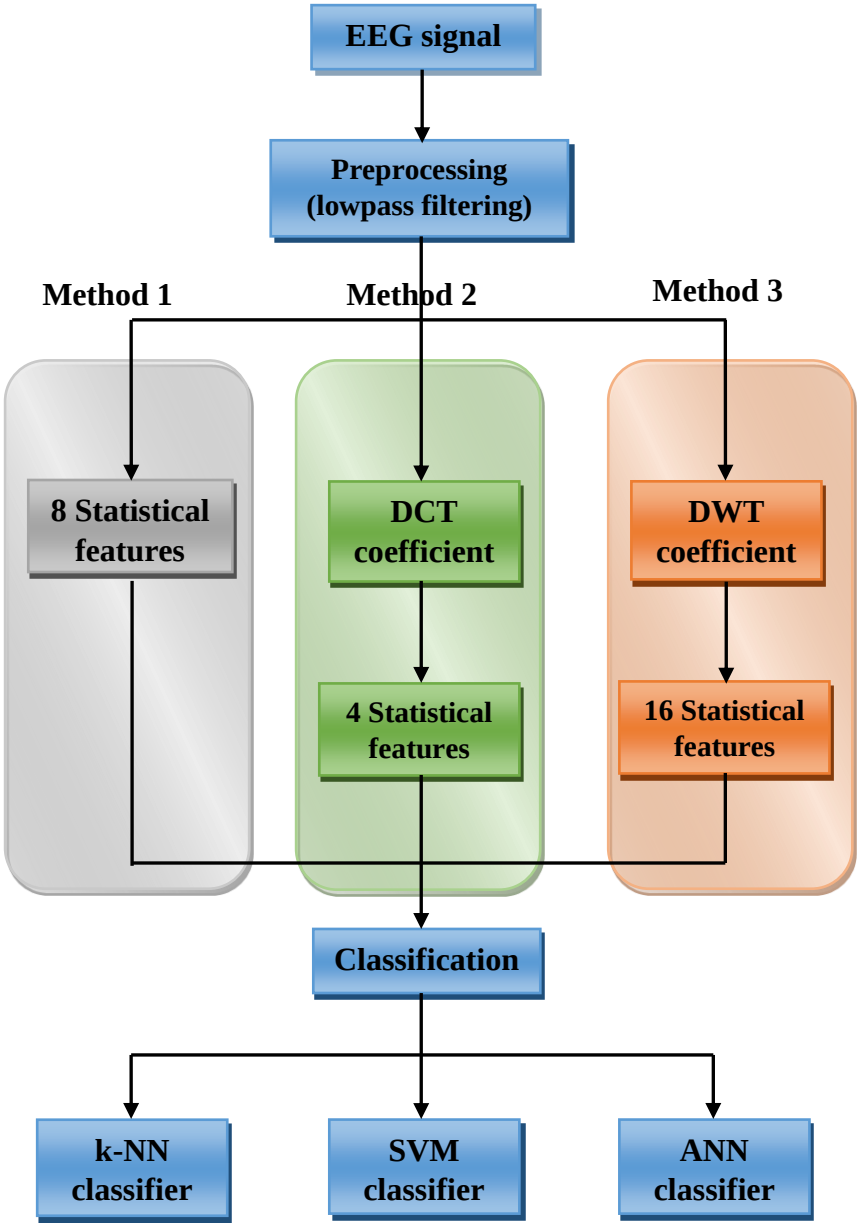


Figure 4.1 The proposed methods for epilepsy and seizure detection.

4.2 Data set description

The used data set was developed by the Department of Epileptology, University of Bonn, Germany. It is made publicly available in [71]. The database consists of five separate

sets denoted set A, B, C, D and E. Each containing 100 single-channel EEG samples of length 23.6s and sampled at 173.6 Hz using 12-bit resolution, resulting in 4097 data points per each signal. The amplitude is in microvolts. All the recordings were made with the same 128-channel amplifier system. Set A and set B were collected from surface EEG recordings of five healthy subjects with eyes open and eyes closed respectively. Sets C, D and E correspond to EEG records of five epileptic patients. The samples in the first two sets are collected during seizure-free intervals (interictal), from the hippocampal formation of the opposite hemisphere of the brain and from within the epileptogenic zone respectively. Set E samples are collected during seizure activity (ictal). The properties of each set are summarized in table 4.1.

Table 4.1 Summary of the main properties of each set within the database.

	Set A	**Set B**	**Set C**	**Set D**	**Set E**
Subject state	Healthy Eyes open	Healthy Eyes closed	Epileptic Interictal	Epileptic Interictal	Epileptic Ictal
Electrode type	Surface	Surface	Intracranial	Intracranial	Intracranial
Electrode placement	International 10-20 system	International 10-20 system	Opposite to epileptogenic zone	Within epileptogenic zone	Within epileptogenic zone

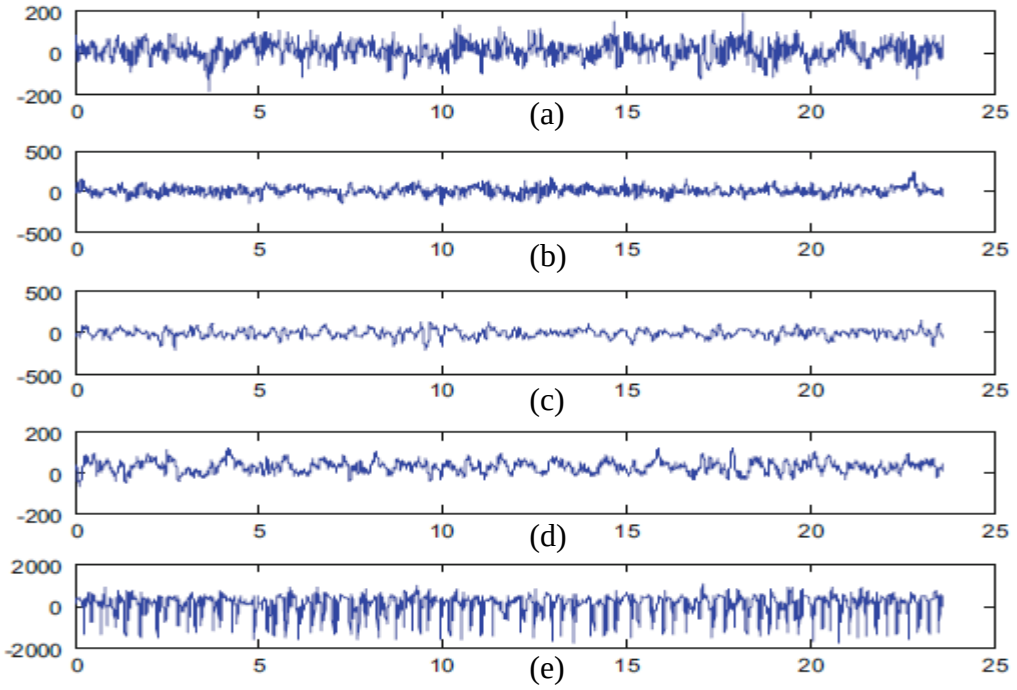


Figure 4.2 Example of an EEG signal from (a) set A (b) set B (c) set C (d) set D (e) set E [72].

Figure 4.2 depicts five samples of the EEG recordings from the five different sets in the Bonn database. The y-axis corresponds to the amplitude in microvolts and the x-axis corresponds to the time in seconds.

4.3 Methodology and procedures

Three methods are proposed for two experiments. In the first experiment, Epilepsy is detected mainly from the interictal intervals and the implemented scenario is healthy vs. epileptic. Therefore, all the samples in the dataset fall in two classes: healthy, for sets A and B, and epileptic for sets C, D and E. In the second experiment, Epilepsy is detected from ictal intervals and the implemented scenario is seizure-free vs. seizure. Since the database has only one set with ictal samples, sets A, B, C and D fall in the first class which is seizure-free (regardless of whether the subject is healthy or epileptic) while set E samples belong to the second class, ictal. *For simplicity, we will refer to the first experiment as Epilepsy detection and the second as seizure detection throughout the whole chapter.*

In order to have good training and validate the results with a test dataset, the Bonn database is quite limited. To tackle this issue an augmentation scheme is proposed. Each EEG signal is divided into 8 signals using a window length of 512 data points with no overlap, as shown in figure 4.3. The resulting samples are treated as independent instances. Therefore, the augmented database has 800 signals per each set, which sums up to a total of 4000 signals.

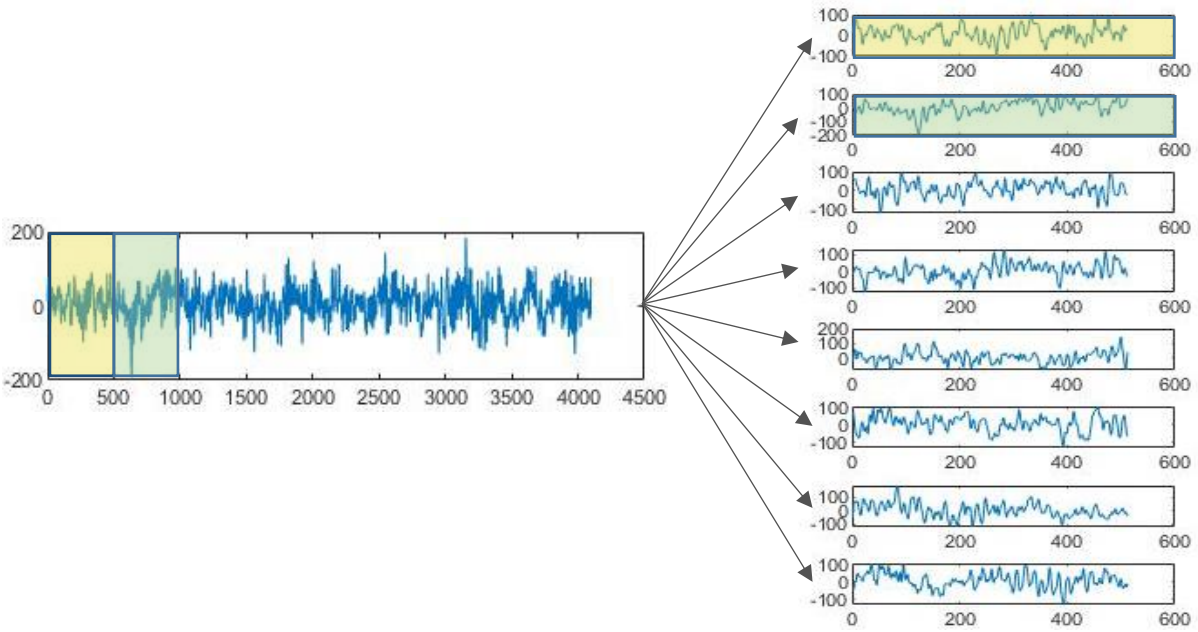


Figure 4.3 Augmentation scheme illustrated in a sample from set A.

4.3.1 Feature extraction

The choice of the right features plays a major role in classification problems. In the first method, eight statistical features are extracted directly from the signal to summarize the relevant information contained in it. Hence, this method relies only on time-domain information. The used statistical features are maximum amplitude, mean, mode, median, standard deviation, first quartile, third quartile and interquartile. The second method relies solely on frequency domain information using the DCT, which is a widely used data compression technique. Since energy is concentrated in low frequencies, as shown in figure 4.4, we keep only the first 150 coefficients (29.3% of the signal after the transformation). Then, we extract four features, which are mean of the absolute value of the coefficients, interquartile, energy and entropy. We will later show that further reduction is possible on the number of input features.

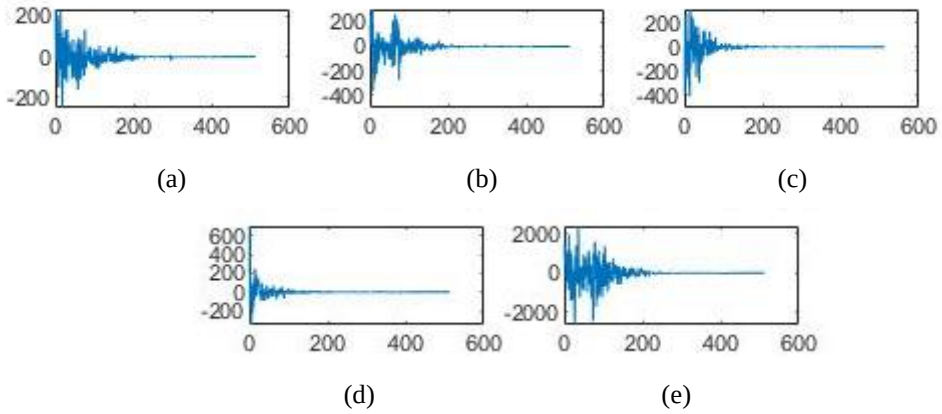


Figure 4.4 DCT of a signal from (a) set A (b) set B (c) set C (d) set D (e) set E.

The third method is based on the DWT, which captures both frequency and location in time information. The first three decomposition levels are tested separately. Figure 4.5 shows the frequency band covered by each level. Both detail and approximation coefficients are taken into account. Figure 4.6 illustrates the plots of detail (in red) and approximation (in blue) coefficients, using the Haar wavelet on a sample from set A. Instead of directly feeding the coefficients to the classifier, we summarize the relevant information in 16 statistical features, 8 for the detail coefficients and 8 for the approximation coefficients. These features are maximum, mean of the absolute value of the coefficients, mode, median, standard deviation, first quartile, third quartile and interquartile.

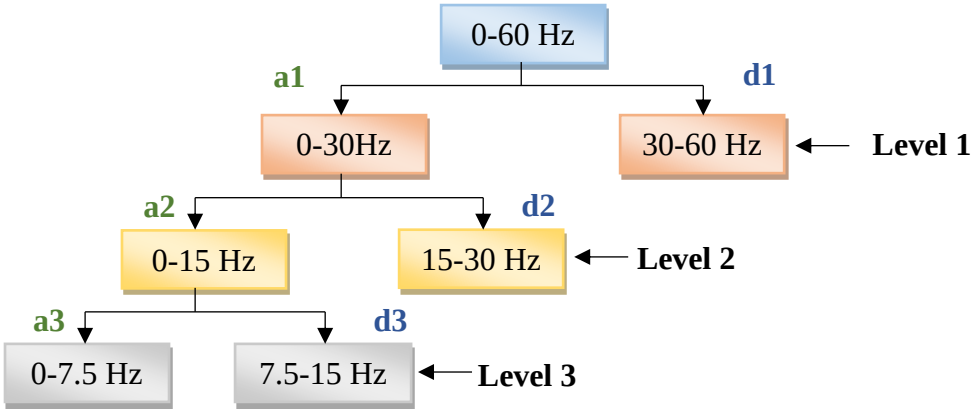
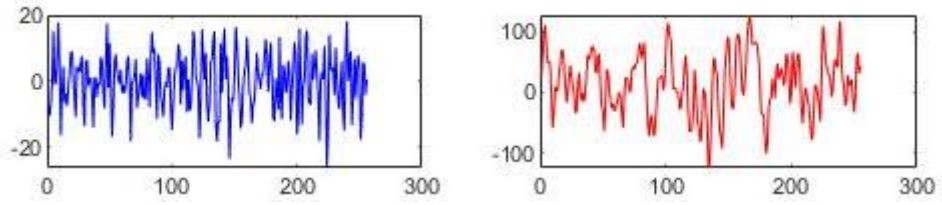
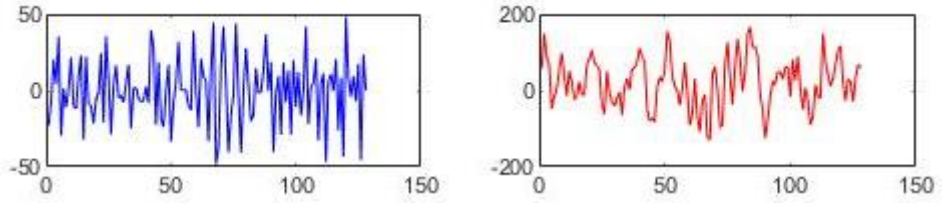


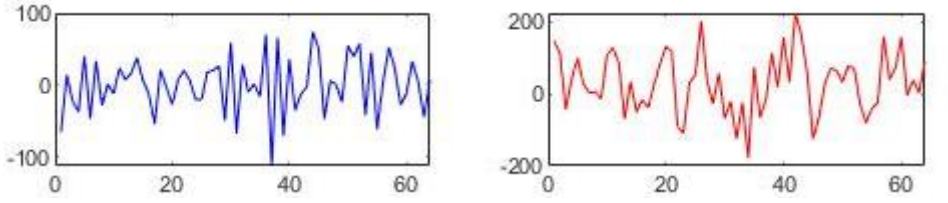
Figure 4.5 The different frequency bands covered by the first three levels of DWT.



(a)



(b)



(c)

Figure 4.6 Discrete Haar wavelet coefficients on a set A signal at (a) level 1 (b) level 2 (c) level 3.

4.3.2 Classification

After extracting the selected features depending on the used method, they are fed to three different classifiers to compare their performances. The first classifier is k-NN, the second is SVM, and the last is ANN. To train both k-NN and SVM models, we used the software Matlab R2018b. The two classifiers are already implemented in the Statistics and Machine Learning Toolbox as the two functions *fitcknn* and *fitcsvm*. To train the ANN classifier, the model was built with Python 3.6. It is made exclusively of dense layers from the Keras library as we are using a simple MLP. The model consists of four hidden layers; the first layer has 30 neurons, while the remaining three were implemented with 20 neurons each. The ReLU activation function was used for the hidden layers, and the sigmoid activation function was chosen for the output layer.

4.3.3 Evaluation

The data is divided into 75% for the training and 25% for testing. The performance metrics used for the evaluation of the model are accuracy, sensitivity, and specificity.
The accuracy (acc) of a classifier is its ability to differentiate between positive and negative cases correctly. Mathematically, it is expressed as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.1}$$

Where *TP* (true positive): the number of cases correctly identified as positive (unhealthy).

TN (true negative): the number of cases correctly identified as negative (healthy).

FP (false positive): the number of cases incorrectly identified as positive.

FN (false negative): the number of cases incorrectly identified as negative.

The sensitivity (sen) of a binary classification model is its ability to determine the positive cases correctly, whereas, the specificity (spe) measures its ability to identify negative cases correctly. They are calculated as follows:

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.2}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.3}$$

4.4 Results and discussions

4.4.1 Experiment 1: Epilepsy detection

In this experiment, the goal is to identify whether a subject has Epilepsy or not mainly from interictal intervals. Several data samples of the Bonn database are tested. First, a pair from the four sets (excluding set E) is taken each time (a healthy set and an epileptic set) resulting in four combinations: A-C, A-D, B-C, and B-D. Then, sets A and B are grouped to form the healthy class while sets C and D form the epileptic class. Finally, set E is added to the latter. For each pair, 1200 samples are used for the training, and 400 samples for the testing. In each train and test dataset the positive and negative cases are equal. The data sample AB-CD is divided into 2400 samples for the training and 800 samples for the testing. Here again, the epileptic portion and the healthy portion are of equal size. The last data sample, which includes the whole database, is divided into 3000 samples for training, from which 1200 are healthy cases and 1800 are epileptic cases, and 1000 samples for testing, where 400 are negative cases and the remaining 600 are positive cases.

4.4.1.a Method 1: Feature extraction using statistical parameters

As mentioned before, the first method is based on the extraction of statistical features directly from the original signal in the time domain. The results are recorded in table 4.2, table 4.3 and table 4.4 for the k-NN classifier, SVM classifier and ANN classifier, respectively.

Table 4.2 The obtained results for Epilepsy detection with the k-NN classifier using the first method (statistical features applied on the original signal).

		A-C	A-D	B-C	B-D	AB-CD	AB-CDE
k = 3	**Acc (%)**	74.75	74.75	67.75	72	66.12	69.8
	Sen (%)	67	69	49	62.5	56.5	67.17
	Spe (%)	82.5	80.5	86.5	81.5	75.75	73.75
k = 5	**Acc (%)**	77	76.25	70	71.25	68	71.2
	Sen (%)	70	67	51	60	55.25	65.67
	Spe (%)	84	85.5	89	82.5	80.75	79.5
k = 8	**Acc (%)**	78.25	76.25	70.5	74.75	68.12	72.3
	Sen (%)	73.5	71	54	67	60.5	62
	Spe (%)	83	81.5	87	82.5	75.75	87.75

Table 4.3 The obtained results for Epilepsy detection with the SVM classifier using the first method (statistical features applied on the original signal).

	A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Acc (%)	82.75	81.75	73	78.75	74.87	75.2
Sen (%)	80.5	71.5	56.5	69	66.25	75
Spe (%)	85	92	89.5	88.5	83.5	75.5

Table 4.4 The obtained results for Epilepsy detection with the ANN classifier using the first method (statistical features applied on the original signal).

	A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Acc (%)	72.5	78.5	71	74.75	69.5	76
Sen (%)	65	66	63.5	74	55	79.33
Spe (%)	80	91	78.5	75	84	75.5

When using the k-NN classifier, changing the parameter k affects the accuracy, such that it increases when we increase the number of nearest neighbors. The average accuracy is 73.36% for k = 8, which makes the k-NN classifier the least performing in this case, followed by ANN with an average accuracy of 73.7%. The SVM classifier has the best performance with an average accuracy of 77.72%. Generally, the pairs with set A as the healthy set give better results than with set B. It is worth noting that the two resting states eyes-open and eyes-closed have different impacts on the brain activity, which results in the observed difference. Mostly, the recorded specificity is higher than the sensitivity. In other words, the three models tend to misclassify the epileptic cases more than the healthy cases. The first method resulted in poor performance. The time-domain information alone is far from enough for Epilepsy detection. Figure 4.7 shows the different features used for four samples from set A, set B, set C and set D. There is no obvious distinction between the healthy and epileptic cases, which would explain the confusion of the classifiers. However, since set E signals are recorded during the seizure, they are distinguishable from the rest as shown in figure 4.8.

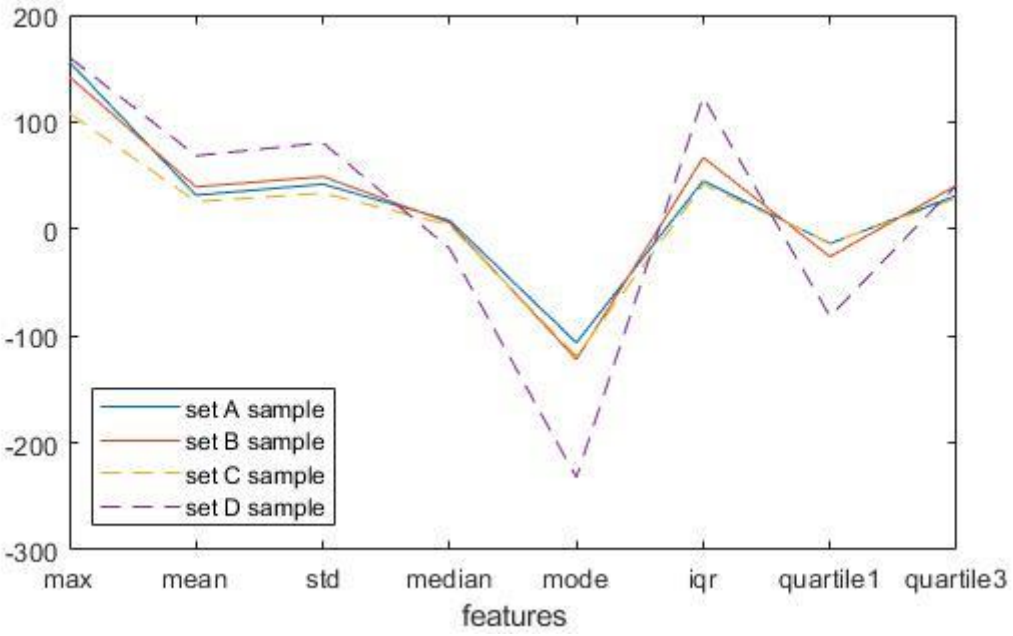


Figure 4.7 An example of four different input signals from sets A, B, C and D used in the training for the first method.

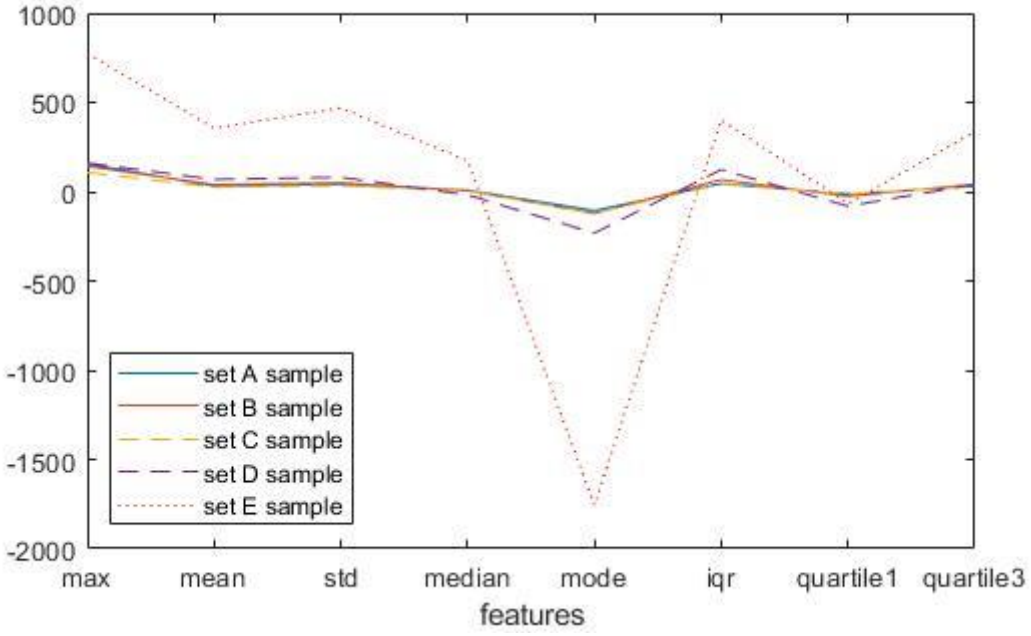


Figure 4.8 An example of different input signals from the five sets used in the training for the first method.

4.4.1.b Method 2: Feature extraction using DCT

Since extracting the statistical features directly from the original signal resulted in a bad performance, we moved to the frequency domain with the DCT to see if that leads to any improvement. Using the four features mentioned in section 4.4.1, the results are recorded in table 4.5 and table 4.6 for the classifiers k-NN and SVM respectively.

Table 4.5 The obtained results for Epilepsy detection with the k-NN classifier using the second method (four statistical features applied on the DCT coefficients).

		A-C	A-D	B-C	B-D	AB-CD	AB-CDE
k = 3	**Acc (%)**	91.5	90.75	81.5	89.75	85.25	88
	Sen (%)	85.5	86.5	65.5	84	78	84.83
	Spe (%)	91.5	95	97.5	95.5	92.5	92.75
k = 5	**Acc (%)**	91.5	92	81	91.5	87	88.8
	Sen (%)	85.5	87.5	64	86.5	80	86
	Spe (%)	97.5	96.5	98	96.5	94	93
k = 8	**Acc (%)**	91.75	92.75	84.25	91	87.87	88.7
	Sen (%)	86	91	70.5	87	82.75	84.33
	Spe (%)	97.5	94.5	98	95	93	95.25

Table 4.6 The obtained results for epilepsy detection with the SVM classifier using the second method (four statistical features applied on the DCT coefficients).

	A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Acc (%)	94.25	92.25	81.5	91.5	87.87	89.2
Sen (%)	90	86.5	64	84	77	84.83
Spe (%)	98.5	98	99	99	98.75	95.75

The best average accuracy with the k-NN classifier, 89.39%, was again achieved with parameter k=8. The SVM model performed barely better with an average accuracy of 89.43%. The performance was especially bad with data sample B-C compared to the other pairs where the accuracy was greater than 90%. The correctly classified cases are not equally distributed over the two classes with both models, as they tend to “favor” the healthy class. The specificity recorded with the SVM classifier was generally greater than 98% (except with data sample AB-CDE), unlike the sensitivity, which was quite low. k-NN was slightly better, as it offers more balance between the two metrics.

To see if there were any redundant features in the input vector, we removed one feature at a time and observed the results. We concluded that the dimensionality could be reduced to half the original one. Both energy and entropy were redundant and therefore removed. The results are shown in table 4.7, table 4.8 and table 4.9 for the classifiers k-NN, SVM and ANN, respectively.

Table 4.7 The obtained results for Epilepsy detection with the k-NN classifier (k=8) using the second method after the dimensionality reduction of the input vector.

	A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Acc (%)	93.5	92	86.75	91.75	88.75	89.1
Sen (%)	90.5	88.5	75	85.5	82.5	84.17
Spe (%)	96.5	95.5	98.5	98	95	96.5

Table 4.8 The obtained results for Epilepsy detection with the SVM classifier using the second method after the dimensionality reduction of the input vector.

	A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Acc (%)	94.25	92.25	84	90.75	87.75	89.3
Sen (%)	90.5	87	69	82.5	77.5	85.33
Spe (%)	98	97.5	99	99	98	95.25

Table 4.9 The obtained results for Epilepsy detection with the ANN classifier using the second method after the dimensionality reduction of the input vector.

	A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Acc (%)	93	92	87.75	88.75	89.37	90.1
Sen (%)	88.5	86	76	78.5	80.75	85
Spe (%)	97.5	98	99.5	99	98	97.75

After reducing the size of the input vector, the best average accuracy recorded is 90.30% with the k-NN classifier (a gain of almost 1%), followed by ANN with an average accuracy of 90.16%, then SVM with an average accuracy of 89.72%. Relying on the frequency domain information with the DCT improved the performance considerably compared to the first method. The gain is 16.94% with k-NN, 16.46% with ANN and 12% with SVM. Nevertheless, the results for some data samples are still not satisfying, especially the sensitivity, which is quite low in most cases.

4.4.1.c Method 3: Feature extraction using DWT

As an attempt to farther improve the performance for Epilepsy detection, we have used a powerful mathematical tool, which is the DWT, to extract the statistical features from the generated approximation and detail coefficients. We have recorded the results for 37 wavelets from six different families, which are Haar, Daubechies, biorthogonal, Coiflet, Symlet and discrete Meyer. We tested the three first decomposition levels separately, but only the best accuracy was recorded with the corresponding level. The complete tables for SVM and k-NN are shown in the appendix, whereas table 4.10 and table 4.11 show only 6 wavelets for which the accuracy was highest with k-NN and SVM classifiers respectively. Table 4.12 refers to the results achieved with ANN.

Table 4.10 The obtained results for epilepsy detection with the k-NN classifier using the third method (extracting statistical features from the DWT coefficients).

Data sample	Wavelet	Level	Acc (%)	Sen (%)	Spe (%)
A-C	Db5	2	92	84.5	99.5
	Db7	1	**93**	**88.5**	**97.5**
	Bior2.4	1	92.5	86	99
	Bior2.6	1	91.75	85	98.5
	Bior5.5	2	91.75	84.5	99
	Coif4	2	92.75	87.5	98
A-D	Db10	3	93	88.5	97.5
	Bior2.4	3	**93.25**	**88**	**98.5**
	Bior3.3	3	93	86.5	99.5
	Bior4.4	3	93.25	88.5	98
	Bior5.5	3	93	89	97
	Sym5	3	92.75	86.5	99
B-C	Db3	3	93.75	87.5	100
	Db5	3	93	87	99
	Db7	3	93	86	100
	Db9	3	93.5	87.5	99.5
	Db10	3	**93.75**	**88**	**99.5**
	Sym3	3	93.75	87.5	100
B-D	Db6	3	97.75	95.5	100
	Db10	3	98	96	100
	Bior4.4	3	97.75	95.5	100
	Bior5.5	3	97.75	97	98.5
	Coif4	3	97.75	96.5	99
	Sym8	3	**98.75**	**97.5**	**100**
AB-CD	Db5	3	91	84	98
	Db7	3	91.37	84.5	98.25
	Db10	3	**92.25**	**86.25**	**98.25**
	Bior6.8	3	91.5	84.5	98.5
	Coif4	3	91.62	85	98.25
	Sym5	3	91.12	83	99.25
AB-CDE	Db3	3	92	88.5	97.25
	Db5	3	91.7	87	98.75
	Db10	3	91.8	89	96
	Coif3	3	91.6	86.83	98.75
	Sym3	3	92	88.5	97.25
	Sym5	3	**92.3**	**88.5**	**98**

Table 4.11 The obtained results for Epilepsy detection with the SVM classifier using the third method (extracting statistical features from the DWT coefficients).

Data sample	Wavelet	Level	Acc (%)	Sen (%)	Spe (%)
A-C	Db2	2	**95.75**	**93**	**98.5**
	Bior1.3	1	94.75	93.5	96
	Bior2.4	1	94.5	94	95
	Bior2.6	1	94.5	94	95
	Sym2	1	95.75	94	97.5
	Sym6	1	94.5	93	96
A-D	Db3	3	93.5	90.5	96.5
	Db5	2	93.75	88.5	99
	Bior2.4	3	93.5	92.5	94.5
	Bior2.8	3	**93.75**	**91**	**96.5**
	Bior5.5	3	93.5	89.5	97.5
	Sym3	3	93.5	90.5	96.5
B-C	Db1	2	84.25	68.5	100
	Db2	3	83.5	67	100
	Db3	3	**84.5**	**69**	**100**
	Sym2	3	83.5	67	100
	Sym3	3	84.5	69	100
	Sym7	3	83.5	67	100
B-D	Db1	2	95.25	93.5	97
	Db10	3	93.75	87.5	100
	Bior1.3	1	95.75	94	97.5
	Bior1.5	1	**96.5**	**94.5**	**98.5**
	Coif4	3	94	88	100
	Sym8	3	93.25	86.5	100
AB-CD	Db1	1	**91**	**84.75**	**97.25**
	Db3	3	88.25	77.75	98.75
	Bior1.3	1	89.37	79.75	97
	Bior1.5	1	89.5	81.25	97.75
	Coif3	3	88	76.75	99.25
	Sym3	3	88.25	77.75	98.75
AB-CDE	Db3	3	**94.6**	**92**	**98.5**
	Db5	3	94.3	90.83	99.5
	Bior1.3	1	93.8	90	99.5
	Sym3	3	94.6	92	98.5
	Sym4	3	93.9	90.83	98.5
	Sym5	3	94.1	91	98.75

Table 4.12 The obtained results for Epilepsy detection with ANN classifier using the third method (extracting statistical features from the DWT coefficients).

Data sample	Wavelet	Level	Acc (%)	Sen (%)	Spe (%)
A-C	Bior2.4	1	90.5	83	98
A-D	Bior2.4	3	93.75	88.5	99
B-C	Db3	3	94.5	90.5	98.5
B-D	Coif4	3	98	96	100
AB-CD	Db10	3	94	90	98
AB-CDE	Db3	3	93.5	91.17	97

We observe from the obtained results that there is no "best wavelet" for EEG data, which would give the highest accuracy for all cases. It depends on both the data sample and the selected classifier. However, the db10 wavelet achieved the best average accuracy of 93.26% with k-NN. SVM was especially sensitive to the change in the training data such that the performance drops drastically with the sample B-C. It is also the least performing classifier with an average accuracy of 92.68%. k-NN was more stable and the least sensitive to data change, wavelet and level change. The average accuracies for the two classifiers, k-NN and ANN were 93.88% and 94.04% respectively. Probably, better results could be obtained with the latter since we tested the model with only one wavelet for each data sample. The choice of the wavelet for ANN was based on the results obtained with the two other classifiers. We choose one with which the accuracy was high for both classifiers. The DWT has indeed improved the overall performance. All the samples have a higher accuracy than 90% (except with SVM). The sensitivity is still lower than the specificity, but considerably high compared to the previous method.

● After carrying on the experiment with the whole EEG signals and deducing that the DWT based method has the best accuracy for Epilepsy detection, we decided to test the performance on the separate EEG rhythms and see whether we can achieve close results with only one rhythm. The rhythms were obtained from filtering the original signal using a second-order Butterworth filter. The wavelet used throughout the whole experiment is db7 (Daubechies order 7). The wavelet choice was not random, it was obtained empirically, but there is no guarantee that this is the best choice. It is worth noting that unlike when using the whole signal, changing the wavelet when dealing with the rhythms separately could lead to very different results (up to 20% difference in the accuracy was observed when testing different wavelets). The used classifiers are SVM and k-NN; however, since the latter achieved better results with all rhythm, which are shown in table 4.13, we included the results obtained with SVM in the appendix.

Table 4.13 The obtained results for Epilepsy detection with the k-NN classifier using the DWT coefficients after decomposing the EEG signal into 5 rhythms.

		A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Delta Rhythm	**Acc (%)**	92.75	93.25	94.25	96.75	93.12	92.9
	Sen (%)	86	87.5	91	94	88.25	90.17
	Spe (%)	99.5	99	97.5	99.5	98	97
Theta Rhythm	**Acc (%)**	87.5	87.25	88.5	91.5	88.12	90.8
	Sen (%)	79	78.5	85.5	89.5	83	88.17
	Spe (%)	96	96	91.5	93.5	93.25	97.75
Alpha Rhythm	**Acc (%)**	76.5	84.25	88.5	91.75	82.12	82.1
	Sen (%)	64	76	78.5	84.5	73	78.5
	Spe (%)	89	92.5	98.5	99	91.25	87.5
Beta Rhythm	**Acc (%)**	78	81.25	84.25	91.25	81.62	83
	Sen (%)	63.5	69	71	83.5	70.5	77
	Spe (%)	92.5	93.5	97.5	99	92.75	92
Gamma Rhythm	**Acc (%)**	80	83.5	88.5	85.5	81.12	83.1
	Sen (%)	71	72.5	78.5	77.5	70.75	78.17
	Spe (%)	89	94.5	98.5	93.5	91.5	90.5

We observe from the obtained results that Epilepsy is better detected in low frequency elements (<8Hz). The best performance was recorded with the Delta rhythm, which has the lowest frequency band (<4Hz), and the highest average accuracy, 93.84%, followed by the theta rhythm (4Hz< frequency <8Hz) with an average accuracy of 88.95%. Then, Alpha, Gamma and Beta rhythms with average accuracies 84.20%, 83.62% and 83.23% respectively. The best accuracy was achieved with data sample B-D, 96.75%, which also has the highest sensitivity and specificity, 94% and 99.5% respectively. Using only the Delta rhythm instead of the whole EEG signal leads to almost the same results, with a loss of only 0.04% in accuracy, a gain of 0.03% in sensitivity and a loss of 0.2% in specificity. Using a different method does not forcibly lead to the same conclusions.

4.4.2 Experiment 2: Seizure detection

This experiment aims to identify epileptic seizures from EEG data. Several samples of the Bonn database are tested. First, we take set E, which represents the ictal class, with one of the remaining four sets each time, resulting in four combinations: A-E, B-E, C-E and D-E.

Then, we use the whole database where sets A, B, C and D form the seizure-free class and set E the ictal class. Table 4.14 shows how the data was divided between training and testing the models.

Table 4.14 Data division to train and test the models for seizure detection.

Data sample	purpose	EEG recordings	Seizure-free cases	Ictal cases
Pairs (A-E, B-E, C-E and D-E)	Training	1200	600	600
	Testing	400	200	200
ABCD-E	Training	3000	2400	600
	Testing	1000	800	200

4.4.2.a Method 1: Feature extraction using statistical parameters

After extracting the features from the original signal in time-domain, the results are recorded in table 4.15 with the k-NN classifier, table 4.16 with the SVM classifier, and table 4.17 with the ANN classifier.

Table 4.15 The obtained results for seizure detection with the k-NN classifier using the first method (statistical features applied on the original signal).

		A-E	B-E	C-E	D-E	ABCD-E
k = 3	**Acc (%)**	99.75	96	97.75	94.25	97.1
	Sen (%)	99.5	93	98.5	94.5	90.5
	Spe (%)	100	99	97	94	98.75
k = 5	**Acc (%)**	99.75	96.25	97.75	95.5	97.2
	Sen (%)	99.5	93	98	96	89.5
	Spe (%)	100	99.5	97.5	95	99.12
k = 8	**Acc (%)**	99.75	95.75	98.25	94.25	96.5
	Sen (%)	99.5	92.5	99	96	90
	Spe (%)	100	99	97.5	92.5	98.12

Table 4.16 The obtained results for seizure detection with the SVM classifier using the first method **(**statistical features applied on the original signal).

	A-E	B-E	C-E	D-E	ABCD-E
Acc (%)	100	95.25	98.5	93.75	95.8
Sen (%)	100	93	99	95	89.5
Spe (%)	100	97.5	98	92.5	97.37

Table 4.17 The obtained results for seizure detection with the ANN classifier using the first method (statistical features applied on the original signal).

	A-E	B-E	C-E	D-E	ABCD-E
Acc (%)	99.75	95.75	98.5	94.75	96.8
Sen (%)	99.5	92.5	99	96	89.5
Spe (%)	100	99	98	93.5	98.62

The performance of the three classifiers was quite good, unlike the obtained results for Epilepsy detection. This is due to the remarkably high peaks in the EEG data, which results from the hyper-activity of the brain during seizure intervals. Figure 4.8 illustrates clearly the big difference in statistical features between set E samples and the other sets. It also justifies why we obtained the lowest accuracy with the data sample D-E. The best set used with set E in the training was set A, which represents the EEG recordings of healthy subjects with eyes open. It resulted in an accuracy of 100% with SVM and 99.75% with both k-NN and ANN. The effect of varying the parameter k in the k-NN model is barely noticeable. The best average accuracy of 97.29%, was recorded with k=5. The least performing classifier was SVM with an average accuracy of 96.66% followed by ANN with an average accuracy of 97.11%. When using the whole database, the sensitivity was especially lower than the specificity compared to the values obtained with the pairs. This is probably due to the unbalance of the positive and negative cases in the training data set. The negative class was 4 times bigger than the positive class, which resulted in lower sensitivity.

4.4.2.b Method 2: Feature extraction using DCT

In the previous experiment, Epilepsy detection, the two features, energy and entropy, were proved redundant in the input vector. However, since we did not want to generalize the observation to this experiment, we observed the results with both 2 and 4 features with the SVM classifier. Once again, the energy and entropy were found to be unnecessary. Therefore, table 4.18, table 4.19, and table 4.20 refer to the obtained results with 2 features, mean and interquartile, with k-NN, SVM and ANN classifiers, respectively.

Table 4.18 The obtained results for seizure detection with the k-NN classifier using the second method (two statistical features applied on the DCT coefficients).

		A-E	B-E	C-E	D-E	ABCD-E
k = 3	**Acc (%)**	100	97.5	96.5	95.25	96.7
	Sen (%)	100	97.5	98.5	97	92.5
	Spe (%)	100	97.5	94.5	93.5	97.75
k = 5	**Acc (%)**	100	97.75	97	96	97.1
	Sen (%)	100	98.5	99.5	98.5	95.5
	Spe (%)	100	97	94.5	93.5	97.5
k = 8	**Acc (%)**	100	97.5	97.25	95.75	97.3
	Sen (%)	100	98.5	99.5	99	97
	Spe (%)	100	96.5	95	92.5	97.37

Table 4.19 The obtained results for seizure detection with the SVM classifier using the second method (two statistical features applied on the DCT coefficients).

	A-E	B-E	C-E	D-E	ABCD-E
Acc (%)	99.75	96.75	98.25	96	96.9
Sen (%)	99.5	96	99	98	92
Spe (%)	100	97.5	97.5	94	98.12

Table 4.20 The obtained results for seizure detection with the ANN classifier using the second method (two statistical features applied on the DCT coefficients).

	A-E	B-E	C-E	D-E	ABCD-E
Acc (%)	99.75	97.25	98.25	96.25	97.5
Sen (%)	100	97	99.5	97.5	96.5
Spe (%)	99.5	97.5	97	95	97.75

Relying on the frequency domain information slightly improved the overall performance. The recorded accuracies for data samples B-E and D-E are higher compared to the previous method. Although, the best data combination is still A-E and the worst is still D-E. The best classifier was ANN with an average accuracy of 97.8% followed by k-NN and SVM with an average accuracy of 97.57% (k=5) and 97.53%, respectively. The main advantage of applying the DCT to the original signal before feature extraction over the previous method is the high sensitivity recorded when using the whole database, such that both sensitivity and specificity are greater than 96% with the best classifier ANN.

4.4.2.c Method 3: Feature extraction using DWT

As in the previous experiment, Epilepsy detection, 37 different wavelets from 6 families were tested with k-NN and SVM. The complete tables are shown in the appendix. Table 4.21 and table 4.22 refer to the obtained results, using the DWT coefficients, with the best 6 performing wavelets in each data sample, with k-NN and SVM, respectively. Table 4.23 refers to the results obtained with the ANN classifier using only a single wavelet per data sample.

Table 4.21 The obtained results for seizure detection with the k-NN classifier using the third method (extracting statistical features from the DWT coefficients).

Data sample	Wavelet	Level	Acc (%)	Sen (%)	Spe (%)
A-E	Db1	3	100	100	100
	Db4	3	100	100	100
	Bior2.2	3	100	100	100
	Coif1	3	100	100	100
	Sym2	3	100	100	100
	Dmey	3	100	100	100
B-E	Db1	2	**97**	**94**	**100**
	Db2	1	97	94	100
	Bior2.2	2	97	94	100
	Bior2.4	3	96.75	94	99.5
	Sym2	1	97	94	100
	Sym4	3	96.75	93.5	100
C-E	Bior2.2	3	99.5	100	99
	Bior2.8	3	99.25	99	99.5
	Bior3.3	2	**99.75**	**99.5**	**100**
	Bior3.7	2	99.5	99	100
	Coif1	3	99.25	99.5	99
	Sym4	2	99.25	99	99.5
D-E	Db1	2	98.25	97.5	99
	Db3	3	98.5	99	98
	Db5	3	98.25	98.5	98
	Coif2	3	98.25	99.5	97
	Sym3	3	98.5	99	98
	Sym5	3	**99**	**99.5**	**98.5**
ABCD-E	Db3	3	97.8	91.5	99.37
	Bior2.2	2	97.9	92.5	99.25
	Bior5.5	2	97.9	91	99.62
	Coif1	1	97.8	91.5	99.37
	Sym3	3	97.8	91.5	99.37
	Sym5	3	**98**	**92**	**99.5**

Table 4.22 The obtained results for seizure detection with the SVM classifier using the third method (extracting statistical features from the DWT coefficients).

Data sample	Wavelet	Level	Acc (%)	Sen (%)	Spe (%)
A-E	Db1	3	100	100	100
	Db5	3	100	100	100
	Bior2.6	3	100	100	100
	Coif2	3	100	100	100
	Sym5	3	100	100	100
	Dmey	3	100	100	100
B-E	Db1	2	97.75	95.5	100
	Db2	3	97.75	96	99.5
	Bior2.4	2	98	96.5	99.5
	Bior2.6	2	**98.25**	**96.5**	**100**
	Coif4	3	97.75	95.5	100
	Sym2	3	97.75	96	99.5
C-E	Bior2.2	1	99.5	100	94
	Bior2.4	1	**99.75**	**100**	**99.5**
	Bior2.6	1	99.5	100	99
	Bior2.8	2	99.5	99.5	99.5
	Bior3.1	3	99.5	100	99
	Coif1	1	99.5	100	99
D-E	Db1	3	96.5	100	93
	Db7	3	96.75	99.5	94
	Bior2.6	3	96.75	99.5	94
	Bior3.1	3	97	99	95
	Coif1	3	**97.25**	**100**	**94.5**
	Coif5	3	96.5	99.5	93.5
ABCD-E	Db1	1	97.5	95.5	98
	Bior2.6	1	97.4	94.5	98.12
	Coif1	1	97.6	95.5	98.12
	Coif2	3	97.6	94.5	98.37
	Sym2	2	97.4	95.5	97.87
	Sym5	2	**98**	**97.5**	**98.12**

Table 4.23 The obtained results for seizure detection with ANN classifier using the third method (extracting statistical features from the DWT coefficients).

Data sample	Wavelet	Level	Acc (%)	Sen (%)	Spe (%)
A-E	Db1	3	100	100	100
B-E	Db1	2	97.25	94.5	100
C-E	Bior3.3	2	98.75	97.5	100
D-E	Coif2	3	97.25	95	99.5
ABCD-E	Sym5	3	98.2	91.5	99.87

All three classifiers have led to perfect accuracy (100%) with data sample A-E. The wavelet choice with the latter is quite irrelevant as can be seen in the complete tables (refer to the appendix). The best performing classifier was k-NN with an average accuracy of 98.75%,

followed by SVM with an average accuracy of 98.65%, then ANN with an average accuracy of 98.29%. Again, it is worth noting that only one wavelet was tested with the ANN classifier for each data sample. Therefore, it is highly possible to record better accuracy with different wavelets, and the order is not final. The DWT based method has resulted in the best performance for seizure detection, such that all accuracies, regardless of the data sample and the classifier, were greater than 97%. However, the bior2.2 wavelet achieved the best average accuracy, 98.48% with k-NN. The lowest sensitivity recorded with the best classifier (k-NN) was 92% when using the whole database. Whereas, the specificity did not drop below 98.5%.For all three methods, it is safe to generalize that for the negative class, using set A instead of set B (healthy sets) and set C instead of set D (epileptic interictal sets) during the training leads to higher accuracy in seizure detection.

● As it was done in the previous experiment, Epilepsy detection, we tested the DWT based method with the separate EEG rhythms to see if we can narrow down the input to only one rhythm instead of the whole signal. The wavelet used is db7, and again, there is no guarantee that this is the best choice. Two classifiers were tested, SVM and k-NN. The former has the best performance with all rhythms except Gamma. Table 4.24 refer to the results obtained with the SVM classifier, whereas the recorded results with k-NN can be seen in the appendix.

Table 4.24 The obtained results for seizure detection with the SVM classifier using the DWT coefficients after decomposing the EEG signal into 5 rhythms.

		A-E	B-E	C-E	D-E	ABCD-E
Delta Rhythm	**Acc (%)**	92.75	93.25	94.25	96.75	92.9
	Sen (%)	86	87.5	91	94	90.17
	Spe (%)	99.5	99	97.5	99.5	97
Theta Rhythm	**Acc (%)**	99.75	99	96.5	95.75	97.9
	Sen (%)	99.5	98.5	96	93.5	91.5
	Spe (%)	100	99.5	97	98	99.5
Alpha Rhythm	**Acc (%)**	100	91.5	98.25	98.5	96
	Sen (%)	100	88.5	99	98	86
	Spe (%)	100	94.5	97.5	99	98.5
Beta Rhythm	**Acc (%)**	98.5	96	98.25	98.75	97.6
	Sen (%)	99	94.5	100	98	94
	Spe (%)	98	97.5	96.5	99.5	98.5
Gamma Rhythm	**Acc (%)**	96.75	88.25	94.5	96	92.3
	Sen (%)	94.5	98	98	97.5	91
	Spe (%)	99	78.5	91	94.5	92.62

Generally, the overall performance was good with all five rhythms. The highest average accuracies were achieved with the Beta and Theta rhythms, 97.82% and 97.78% respectively, followed by Alpha with an average accuracy of 96.85%. Then, Delta and Gamma rhythms with average accuracies 95.86% and 93.56%, respectively. The detection of epileptic seizures is higher in the frequency band 4 Hz to 30 Hz. The main difference between the results recorded with Theta and Beta rhythms is that higher accuracies (≥99%) were achieved with the Theta rhythm with the healthy sets (A and B) whereas, the results achieved with the interictal sets (C and D) were better with the Beta rhythm (≥98.25%). Also, the latter has the best average sensitivity, 97.1%, which is 1.3% higher than the sensitivity recorded with the Theta rhythm. However, the average specificity of the latter, 98.8% is 0.8% greater than the recorded average specificity with the Beta rhythm. The results achieved with the Beta rhythm are very close to those achieved with the whole signal. There is a loss of 0.93% in accuracy, a gain of 0.1% in sensitivity, and a loss of 1.6% in specificity. Here again, the drawn conclusions concern only this method. The fact that epileptic seizures were best detected with the Beta rhythm cannot be generalized to other researches with different methods.

4.5 Summary

In this chapter, we presented three methods for two types of problems concerning Epilepsy. The first one is the detection of the disease during seizure-free intervals from EEG data. The second is the identification of the epileptic seizures from the same data. The difference between the presented methods lies in the features extraction stage. In the first method, we directly used the original signal to extract 8 statistical features. In the second and third methods, an extra step is added. In the former, we first obtained the DCT coefficients then summarized the relevant information in 2 features, whereas in the latter, we used the DWT transformation on the signal then we extracted 16 features. We preferred to perform the classification with more than one model. Hence, we used three classifiers k-NN, SVM, and ANN. Several data samples were tested.

For Epilepsy detection, the first method was proved to be the worst with the best accuracy achieved, 82.75%, for the A-C data sample. But, in most cases, the accuracy was less than 80%. The second method, based on the DCT, performed better. The accuracy was greater than 90% for three data samples and greater than 80% for the remaining three. The best overall performance was achieved with the last method based on the DWT. For all data samples, with the k-NN classifier, the minimum accuracy recorded was 92.25%, the minimum sensitivity was 86.25%, and the minimum specificity was 97.5%.

For seizure detection, all three methods had a good performance. Although, the order was the same as in the first experiment. The least average accuracy recorded was 97.29% using the first method (with the k-NN classifier). The DCT slightly improved the results with an average gain of 0.51% in the accuracy (with the ANN classifier). The best performance was recorded with the DWT based method, where the average accuracy was 98.75%, the average sensitivity was 97%, and the average specificity was 99.6% (with the k-NN classifier).

The last step in both experiments was to test the DWT based method on the five rhythms extracted from the EEG signal. We observed that for Epilepsy detection, almost the same performance could be achieved from only the Delta rhythm. Whereas for seizure detection, very close results to those recorded with the whole signal were achieved from the Beta rhythm.

General conclusion

The EEG test gives information about the electrical activity carried out in the brain. It is the most suitable test for Epilepsy diagnosis since epileptic seizures are characterized by the abnormal brain activity and the unnaturally high spikes of voltage recorded during seizure.

Many researches were carried out in order to automatize the diagnosis using machine learning. Most of them are based on seizure detection for Epilepsy diagnosis. Our contributions in this study are that we worked on the diagnosis during both ictal (during seizure) and interictal (seizure-free) activities in two different experiments; we have used three techniques for the feature extraction stage and three different classifiers to compare their performance, and we decomposed the EEG signal into five rhythms to deduce the best rhythm for the diagnosis. The first technique is based on the time domain information only, the second on the frequency domain information only and the third is based on both.

Extracting statistical features directly form the time domain signal was the least performing technique especially during interictal intervals. Using the DCT on the signal then extracting statistical features from the coefficients improved considerably the performance compared to the previous technique. As a last method, we used a powerful analysis tool in the feature extraction stage, which is the DWT. The best performance was recorded with this technique. However, the experimental results showed that the choice of the mother wavelet, the order and the level of decomposition might be very difficult and no prior assumption over what is the best choice may be made before carrying out the experiment. In the classification stage, we used three different classifiers with each method, k-NN, SVM and ANN. With the DWT based method, k-NN had a better overall performance than SVM and was more stable to the wavelet, order and level changes.

The last step in our study was to separate the five rhythms from the EEG signals by filtering to see if we could use only one rhythm as input before the feature extraction stage instead of the whole signal. The results showed that the Delta rhythm, which has the lowest frequency band is enough for Epilepsy detection from interictal intervals. Whereas, the Beta rhythm had the best performance among the five rhythms for seizure detection. However, these findings do not go beyond the database used which is the *Bonn* database with an augmentation scheme, and the method used which is the DWT based method.

For further work, we suggest:

- The Bonn database used throughout this study is a general database. We would like to obtain a more specific dataset where we will be able to study the effect of gender and age on the EEG recordings of epileptic patients.

- The statistical features extracted from the DWT were not proven to be all necessary. Therefore, we suggest using a dimensionality reduction technique such as PCA, in order to get rid of any possible redundant features.

- The deep learning branch of machine learning and the more sophisticated neural networks were not explored in this study. The obtained results might be further improved using the prestigious Recurrent Neural Network- Long Short Term Memory (RNN-LSTM) model which allows the neural network to retain memory through its feedback connections and smart neurons. The feature extraction stage might be skipped altogether and the signals can be directly fed to the classifier as LSTM is a very suited tool for sequence problems.

- In this study and all the previous research carried out about the current topic, the seizures are detected after their occurrence. In the future, it will be interesting to investigate these findings in order to build a forecasting model able to detect the seizures before their occurrence.

Appendix

Table 1. The obtained results for Epilepsy detection with the k-NN classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples A-C, A-D and B-C.

	A-C				A-D				B-C			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**	**level**	**acc**	**sen**	**spe**
Db1	1	89.5	81	98	2	90.75	83	98.5	3	92.5	86	99
Db2	1	90.75	83	98.5	2	92.75	87.5	98	3	91.25	82.5	100
Db3	1	90	82	98	3	91.5	84.5	98.5	3	93.75	87.5	100
Db4	2	91	83	99	2	91.75	84	99.5	3	92.25	84.5	100
Db5	2	92	84.5	99.5	3	92.5	86.5	98.5	3	93	87	99
Db6	2	91.25	84	98.5	3	91	85.5	96.5	3	91.5	83.5	99.5
Db7	1	93	88.5	97.5	3	92	86	98	3	93	86	100
Db8	2	88.75	80	97.5	3	91.5	86	97	3	91.25	82.5	100
Db9	2	90.5	81.5	99.5	3	91	85	97	3	93.5	87.5	99.5
Db10	2	90.75	83	98.5	3	93	88.5	97.5	3	93.75	88	99.5
Bior1.3	1	89.75	80	99.5	1	89.75	80.5	99	1	91.75	84.5	99
Bior1.5	1	89	79.5	98.5	2	90.25	82	98.5	2	91.75	83.5	100
Bior2.2	1	91.25	84	98.5	3	91.25	83.5	99	3	92.25	84.5	100
Bior2.4	1	92.5	86	99	3	93.25	88	98.5	3	92	84.5	99.5
Bior2.6	1	91.75	85	98.5	3	91.75	86	97.5	3	91.25	82.5	100
Bior2.8	1	91	84	98	3	92.5	86	99	3	89.5	81	98
Bior3.1	3	91.25	86.5	96	3	91.25	85	97.5	3	86.25	74	98.5
Bior3.3	1	89	80.5	97.5	3	93	86.5	99.5	3	87.75	77	98.5
Bior3.5	1	89.75	81.5	98	3	91.25	83.5	99	3	85.25	72.5	98
Bior3.7	1	89.75	82.5	97	3	90.5	83	98	3	87.75	77	98.5
Bior3.9	1	89.25	81	97.5	3	90	82	98	3	86	75	97
Bior4.4	3	91.25	84	98.5	3	93.25	88.5	98	3	92.25	86	98.5
Bior5.5	2	91.75	84.5	99	3	93	89	97	3	92.25	86	98.5
Bior6.8	3	90.25	82	98.5	3	92.25	86	98.5	3	89.75	80.5	99
Coif1	1	91	83.5	98.5	1	91.75	87	96.5	3	90.75	82.5	99
Coif2	1	90.75	84.5	97	2	92	84.5	99.5	3	91.5	83.5	99.5
Coif3	2	91.25	83	99.5	3	90.5	84.5	96.5	3	91.75	83.5	100
Coif4	2	92.75	87.5	98	3	90.75	85	96.5	3	90.75	82	99.5
Coif5	1	90.75	86.5	95	3	92.25	86	98.5	3	92	84	100
Sym2	1	90.75	83	98.5	2	92.75	87.5	98	3	91.25	82.5	100
Sym3	1	90	82	98	3	91.5	84.5	98.5	3	93.75	87.5	100
Sym4	1	90.5	84	97	3	92	86.5	97.5	3	91.5	83.5	99.5
Sym5	2	90.75	82	99.5	3	92.75	86.5	99	3	91.75	83.5	100
Sym6	1	89.75	83	96.5	3	91	85	97	3	91	83	99
Sym7	2	90.5	82	99	3	91	83.5	98.5	3	90.75	81.5	100
Sym8	1	91.25	87.5	95	3	91.5	85	98	3	90.75	81.5	100
Dmey	1	91.5	85.5	97.5	3	90.5	85.5	95.5	3	90.25	80.5	100

Table 2. The obtained results for Epilepsy detection with the k-NN classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples B-D, AB-CD and AB-CDE.

	B-D				AB-CD				AB-CDE			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**	**level**	**acc**	**sen**	**spe**
Db1	2	96	82	100	2	90	81.75	98.75	2	90.2	85.33	97.5
Db2	3	97.25	94.5	100	2	89.87	81	98.75	3	90.7	86.33	97.25
Db3	3	95.75	92.5	99	3	90.75	83.75	97.75	3	92	88.5	97.25
Db4	3	97	95	99	3	90.25	83.5	97	3	91	86.67	97.5
Db5	3	96.75	94	99.5	3	91	84	98	3	91.7	87	98.75
Db6	3	97.75	95.5	100	3	90.5	83.75	97.25	3	91.3	86.67	98.25
Db7	3	96.25	93	99.5	3	91.37	84.5	98.25	3	91.5	87	98.25
Db8	3	97	94.5	99.5	3	91	83	99	3	91.6	87.33	98
Db9	3	96.75	94.5	99	3	89.37	82	96.75	3	91.1	86.83	97.5
Db10	3	98	96	100	3	92.25	86.25	98.25	3	91.8	89	96
Bior1.3	3	95.25	92.5	98	1	89.37	79.75	99	1	89.6	83.83	98.25
Bior1.5	2	95.75	92.5	99	2	89.62	80.25	99	2	89.2	84.33	96.5
Bior2.2	3	95	90	100	3	89.5	80.25	98.75	3	89.3	83	98.75
Bior2.4	3	95.75	92.5	99	3	89.62	81	98.25	3	90.5	85.67	97.75
Bior2.6	3	94.75	90	99.5	3	89	79.75	98.25	3	89.3	84	97.25
Bior2.8	3	96.75	93.5	100	3	90.37	82.25	98.5	3	89.7	83.67	98.75
Bior3.1	3	95.25	91	99.5	3	89.12	79.75	98.5	3	87	80.83	96.25
Bior3.3	3	91.75	84.5	99	3	87.25	77.5	97	3	86.1	79.5	96
Bior3.5	3	93.75	88.5	99	3	88	79	97	3	86.4	80.83	94.75
Bior3.7	3	94	88	100	3	88.87	78.75	99	3	86.5	81.33	94.25
Bior3.9	3	93.25	88.5	98	3	87.37	77.75	97	3	86.9	82.17	94
Bior4.4	3	97.75	95.5	100	3	90.12	82.75	97.5	3	90.9	86.5	97.5
Bior5.5	3	97.75	97	98.5	3	90.75	83.5	98	3	90.2	85.67	97
Bior6.8	3	97.75	95.5	100	3	91.5	84.5	98.5	3	90.6	85.17	98.75
Coif1	3	95.75	92	99.5	3	90	82.75	97.25	3	90.4	86	97
Coif2	3	97.25	94.5	100	3	90.62	83	98.25	3	91	86.17	98.25
Coif3	3	97	94	100	3	89.75	81.5	98	3	91.6	86.83	98.75
Coif4	3	97.75	96.5	99	3	91.62	85	98.25	3	91.2	87	97.5
Coif5	3	97.25	95	99.5	3	90.75	83.5	98	3	91.2	87.17	97.25
Sym2	3	97.25	94.5	100	2	89.87	81	98.75	3	90.7	86.33	97.25
Sym3	3	95.75	92.5	99	3	90.75	83.75	97.75	3	92	88.5	97.25
Sym4	3	96.75	93.5	100	3	90	82.5	97.5	3	91.2	87.17	97.25
Sym5	3	97	95	99	3	91.12	83	99.25	3	92.3	88.5	98
Sym6	3	97.5	95.5	99.5	3	90.87	84.25	97.5	3	91.3	86.83	98
Sym7	3	96.25	92.5	100	3	90.5	82	99	3	91.5	86.83	98.5
Sym8	3	98.75	97.5	100	3	90.87	84.25	97.5	3	90.2	86.17	96.25
Dmey	3	97	94	100	3	90.12	83.5	96.75	3	91.1	87	97.25

Table 3. The obtained results for Epilepsy detection with the SVM classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples A-C, A-D and B-C.

	A-C				A-D				B-C			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**	**level**	**acc**	**sen**	**spe**
Db1	1	94.5	92.5	96.5	1	92	86	98	2	84.25	68.5	100
Db2	2	95.75	93	98.5	1	93	88	98	3	83.5	67	100
Db3	1	94.25	92	96.5	3	93.5	90.5	96.5	3	84.5	69	100
Db4	1	93.5	91	96	2	91.25	87	95.5	3	81	62	100
Db5	1	93.75	91	96.5	2	93.75	88.5	99	3	83.25	66.5	100
Db6	1	93.75	93	94.5	2	91.5	84	99	3	81.25	62.5	100
Db7	1	93.5	93	94	1	91.5	85.5	97.5	3	83.25	66.5	100
Db8	1	92.25	90.5	94	2	92.25	87.5	97	3	82	64	100
Db9	1	93	93	93	3	92	89	95	3	82.75	65.5	100
Db10	1	92.5	92.5	92.5	3	92.5	90.5	94.5	3	82	64	100
Bior1.3	1	94.75	93.5	96	1	91	85	97	1	80.5	61	100
Bior1.5	1	93.5	92.5	94.5	1	92	88.5	95.5	2	81.75	63.5	100
Bior2.2	1	94	93	95	2	93	89	97	3	81.5	63	100
Bior2.4	1	94.5	94	95	3	93.5	92.5	94.5	3	81.5	63	100
Bior2.6	1	94.5	94	95	2	93.25	89.5	97	3	79.5	59	100
Bior2.8	1	94.5	94	95	3	93.75	91	96.5	3	80.75	61.5	100
Bior3.1	1	92.5	89.5	95.5	3	91.25	87.5	95	3	80.25	60.5	100
Bior3.3	1	91	86	96	3	91.75	89	94.5	3	78.75	57.5	100
Bior3.5	1	90.25	84	96.5	3	91.75	88.5	95	3	76.75	54.5	99
Bior3.7	1	91.75	87	96.5	3	92.25	90	94.5	3	77.75	55.5	100
Bior3.9	1	92.25	89	95.5	3	90.5	86.5	94.5	3	79	58	100
Bior4.4	1	93.5	91.5	95.5	3	92.25	92	92.5	3	82.5	65.5	99.5
Bior5.5	2	92.5	86	99	3	93.5	89.5	97.5	3	82.75	65.5	100
Bior6.8	1	93	93.5	92.5	3	92.5	90	95	3	82	64	100
Coif1	1	93.25	91.5	95	2	92.75	87.5	98	3	83.25	66.5	100
Coif2	1	94.25	92.5	96	2	92.25	87	97.5	3	82.25	64.5	100
Coif3	1	94	94.5	93.5	3	91.75	86.5	97	3	82	64	100
Coif4	1	94	93	95	3	91.25	89	93.5	3	80.75	61.5	100
Coif5	1	93.25	93	93.5	3	92	88.5	95.5	3	82.5	65	100
Sym2	1	95.75	94	97.5	1	93	88	98	3	83.5	67	100
Sym3	1	94.25	92	96.5	3	93.5	90.5	96.5	3	84.5	69	100
Sym4	1	93.25	89.5	97	2	93.25	88	98.5	3	83	66	100
Sym5	1	94.25	92.5	96	3	92.25	89.5	95	3	83.25	66.5	100
Sym6	1	94.5	93	96	3	93	91	95	3	81.5	63.5	99.5
Sym7	1	93.5	93.5	93.5	3	93	89.5	96.5	3	83.5	67	100
Sym8	1	92.5	92.5	92.5	3	91.75	90	93.5	3	80.75	61.5	100
Dmey	1	93	92	94	1	92.5	88.5	96.5	3	82	64	100

Table 4. The obtained results for Epilepsy detection with the SVM classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples B-D, AB-CD and AB-CDE.

	B-D				AB-CD				AB-CDE			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**	**level**	**acc**	**sen**	**spe**
Db1	2	95.25	93.5	97	1	91	84.75	97.25	1	93.7	90	99.25
Db2	3	93	86.5	99.5	2	87.37	75.5	99.25	3	92.9	89.83	97.5
Db3	3	91	82	100	3	88.25	77.75	98.75	3	94.6	92	98.5
Db4	3	92.75	86	99.5	3	87.87	77	98.75	3	93.3	90	98.25
Db5	3	90.5	81	100	3	87.62	75.75	99.5	3	94.3	90.83	99.5
Db6	3	92.25	84.5	100	3	87.62	76.25	99	3	93.7	90.33	98.75
Db7	3	92	84	100	3	87.37	76	98.75	3	93.3	90.5	97.5
Db8	3	91.25	83	99.5	3	87.12	75.25	99	3	93.4	90	98.5
Db9	3	92.25	85	99.5	3	87	76	98	3	92.7	89.5	97.5
Db10	3	93.75	87.5	100	3	87.87	77.5	98.25	3	93.1	89.67	98.25
Bior1.3	1	95.75	94	97.5	1	89.37	9.75	97	1	93.8	90	99.5
Bior1.5	1	96.5	94.5	98.5	1	89.5	81.25	97.75	1	93.1	89.17	99
Bior2.2	2	89.5	79	100	1	85.75	73.25	98.25	3	92.2	87.83	98.75
Bior2.4	3	90.5	82	99	3	86.37	73.25	99.5	3	92.8	89.33	98
Bior2.6	1	90	80	100	1	85. 5	70.25	98.75	3	91.3	86.83	98
Bior2.8	3	91	82	100	1	86.12	74	98.25	3	91.9	87	99.25
Bior3.1	3	89.5	79	100	3	83.5	67.5	99.5	3	89.4	84.33	97
Bior3.3	3	86.75	74	99.5	3	82	64.25	99.75	3	89.4	83.83	97.75
Bior3.5	3	86.75	74	99.5	1	81.62	66.5	96.75	3	89	84.33	96
Bior3.7	3	87.25	74.5	100	3	82.12	64.75	99.5	3	89.6	85.33	96
Bior3.9	3	87.75	76	99.5	3	83.12	67	99.25	3	90.1	86.17	96
Bior4.4	3	92.5	85	100	3	85.87	73.25	98.5	3	93.5	90.17	98.5
Bior5.5	3	93	86.5	100	3	87.5	76	99	3	93.4	90.33	98
Bior6.8	3	91.5	83	100	3	87.12	74.75	99.5	3	92.8	89	98.5
Coif1	2	92.5	85	100	2	87.5	75. 5	99.5	3	93.3	89.83	98.5
Coif2	3	91.75	84	99.5	3	87.25	75.5	99	3	92.8	89	98.5
Coif3	3	92.5	85	100	3	88	76.75	99.25	3	93.1	89.17	99
Coif4	3	94	88	100	3	86.75	75	98.5	3	92.9	90.33	96.75
Coif5	3	92.5	85	100	3	87.75	76.5	99	3	93	89.5	98.25
Sym2	3	93	86.5	99.5	2	87.37	75.5	99.25	3	92.9	89.93	97.5
Sym3	3	91	82	100	3	88.25	77.75	98.75	3	94.6	92	98.5
Sym4	3	92	84	100	3	87.87	76	99.75	3	93.9	90.83	98.5
Sym5	3	92.75	85.5	100	3	87.12	75	99.25	3	94.1	91	98.75
Sym6	3	92.25	85.5	99	3	86.5	74.75	98.25	3	93.8	90.83	98.25
Sym7	3	92	84	100	3	87.25	74.75	99.75	3	93.5	90.33	98.25
Sym8	3	93.25	86.5	100	3	87.2	75.5	98.75	3	92.7	89.5	97.5
Dmey	3	91.5	83	100	3	86.62	74.5	98.75	3	93.1	89.67	98.25

Table 5. The obtained results for seizure detection with the k-NN classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples A-E, B-E and C-E.

	A-E				B-E				C-E			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**	**level**	**acc**	**sen**	**spe**
Db1	3	100	100	100	2	97	94	100	3	99	99.5	98.5
Db2	3	100	100	100	1	97	94	100	2	99	99.5	98.5
Db3	3	100	100	100	2	96.5	93	100	2	98.5	98.5	98.5
Db4	3	100	100	100	1	96	92.5	99.5	2	98.5	97.5	99.5
Db5	2	100	100	100	3	96	92	100	3	98.75	99.5	98
Db6	3	99.75	99.5	100	3	96.25	92.5	100	3	98.5	99.5	97.5
Db7	3	99.75	99.5	100	3	96.75	93.5	100	1	98.5	97.5	99.5
Db8	3	99.75	99.5	100	3	95.75	91.5	100	1	98.25	97	99.5
Db9	3	99.75	99.5	100	3	95.5	91	100	2	98.75	99	98.5
Db10	3	100	100	100	3	96.5	93	100	1	98.25	96.5	100
Bior1.3	2	100	100	100	1	95.75	91.5	100	2	98.5	99	98
Bior1.5	3	99.75	99.5	100	1	96	92	100	1	98.5	97.5	99.5
Bior2.2	3	100	100	100	2	97	94	100	3	99.5	100	99
Bior2.4	3	100	100	100	3	96.75	94	99.5	3	99	99.5	98.5
Bior2.6	2	100	100	100	2	96.75	93.5	100	2	99	98	100
Bior2.8	3	100	100	100	2	96.25	92.5	100	3	99.25	99	99.5
Bior3.1	3	100	100	100	1	96.25	93	99.5	3	98.5	98.5	98.5
Bior3.3	3	99.75	99.5	100	2	95.75	92.5	99	2	99.75	99.5	100
Bior3.5	3	100	100	100	2	95.5	92	99	1	98.5	97	100
Bior3.7	3	100	100	100	2	95.5	92	99	2	99.5	99	100
Bior3.9	3	99.5	99	100	2	95	91.5	98.5	2	98.5	97	100
Bior4.4	2	100	100	100	3	96.75	93.5	100	2	98.75	98	99.5
Bior5.5	3	99.5	99	100	1	96.25	93.5	99	1	98.25	96.5	100
Bior6.8	3	99.75	99.5	100	3	96.25	92.5	100	2	98.25	97	99.5
Coif1	3	100	100	100	2	96.5	93	100	3	99.25	99.5	99
Coif2	2	100	100	100	2	96.5	93.5	99.5	2	98.75	98	98.5
Coif3	3	99.75	99.5	100	2	96.5	93	100	3	99	99.5	98.5
Coif4	3	99.75	99.5	100	2	96.5	93	100	2	98.5	97.5	99.5
Coif5	3	99.75	99.5	100	3	96.5	93	100	3	98.75	99.5	98
Sym2	3	100	100	100	1	97	94	100	2	99	99.5	98.5
Sym3	3	100	100	100	2	96.5	93	100	2	98.5	98.5	98.5
Sym4	2	100	100	100	3	96.75	93.5	100	2	99.25	99	99.5
Sym5	3	100	100	100	2	96.5	93	100	3	98.5	99	98
Sym6	3	100	100	100	3	96.75	93.5	100	3	98.75	99.5	98
Sym7	3	99.75	99.5	100	3	96.75	93.5	100	3	98.75	99	98.5
Sym8	3	99.75	99.5	100	3	96.5	93	100	1	98.75	97.5	100
Dmey	3	100	100	100	3	96.5	93	100	3	98	99	97

Table 6. The obtained results for seizure detection with the k-NN classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples D-E and ABCD-E.

	D-E				ABCD-E			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**
Db1	2	98.25	97.5	99	2	97.5	90.5	99.25
Db2	1	97	98	96	2	97.4	90	99.25
Db3	3	98.5	99	98	3	97.8	91.5	99.37
Db4	3	97.25	96.5	98	1	97.2	91	98.75
Db5	3	98.25	98.5	98	3	97.7	90	99.62
Db6	3	96.5	98	95	3	97.4	90.5	99.12
Db7	3	97.75	98	97.5	3	97.3	90.5	99
Db8	2	97.75	98	97.5	2	97.3	89.5	99.25
Db9	3	98	99.5	96.5	2	97.5	91	99.12
Db10	3	97.25	98.5	96	2	97.4	90.5	99.12
Bior1.3	3	96.75	96	97.5	3	97.2	88.5	99.37
Bior1.5	1	97.5	96.5	98.5	2	97.4	89.5	99.37
Bior2.2	3	98	98	98	2	97.9	92.5	99.25
Bior2.4	3	97.75	98.5	97	3	97.7	92.5	99
Bior2.6	3	97.25	98	96.5	2	97.7	92	99.12
Bior2.8	2	97.5	97	98	1	97.4	89.5	99.37
Bior3.1	3	97.25	98	96.5	2	97.2	90	99
Bior3.3	3	97.5	97.5	97.5	1	97.1	89.5	99
Bior3.5	1	98	99	97	1	97.2	90	99
Bior3.7	2	97	97.5	96.5	1	97.1	90.5	98.75
Bior3.9	3	97	98	96	2	97.2	89.5	99.12
Bior4.4	2	97.25	96.5	98	2	97.6	90	99.5
Bior5.5	2	97.5	97	98	2	97.9	91	99.62
Bior6.8	2	98	98	98	3	97.6	91	99.25
Coif1	3	98	98	98	1	97.8	91.5	99.37
Coif2	3	98.25	99.5	97	3	97.7	91.5	99.25
Coif3	3	97.5	99	96	3	97.5	91	99.12
Coif4	3	97.5	98	97	2	97.5	90	99.37
Coif5	3	97	98	96	2	97.3	90	99.12
Sym2	1	97	98	96	2	97.4	90	99.25
Sym3	3	98.5	99	98	3	97.8	91.5	99.37
Sym4	2	97.75	98.5	97	3	97.4	91	99
Sym5	3	99	99.5	98.5	3	98	92	99.5
Sym6	2	97.5	97.5	97.5	3	97.7	92.5	99
Sym7	3	97.75	99	96.5	3	97.6	91	99.25
Sym8	3	97.5	98	97	2	97.7	92	99.12
Dmey	3	97.5	98.5	96.5	3	97.3	91	98.87

Table 7. The obtained results for seizure detection with the SVM classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples A-E, B-E and C-E.

	A-E				B-E				C-E			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**	**level**	**acc**	**sen**	**spe**
Db1	3	100	100	100	2	97.75	95.5	100	1	99	100	98
Db2	3	100	100	100	3	97.75	96	99.5	1	99	100	98
Db3	3	100	100	100	3	97.25	95	99.5	2	99.25	100	98.5
Db4	3	100	100	100	3	97.25	95	99.5	2	99	99.5	98.5
Db5	3	100	100	100	2	97	95	99	2	98	99	97
Db6	3	100	100	100	3	96.75	94	99.5	2	98.25	100	96.5
Db7	3	100	100	100	3	97.5	95.5	99.5	1	97.75	98.5	97
Db8	3	100	100	100	3	97.25	94.5	100	2	98.25	99	97.5
Db9	3	100	100	100	3	96.75	94.5	99	2	97.5	99.5	95.5
Db10	3	100	100	100	3	97.5	95	100	1	98.25	98.5	98
Bior1.3	3	100	100	100	2	97.25	94.5	100	1	98.5	99.5	97.5
Bior1.5	3	100	100	100	3	97	95	99	1	98.25	99.5	97
Bior2.2	3	100	100	100	2	97.75	95.5	100	1	99.5	100	94
Bior2.4	3	100	100	100	2	98	96.5	99.5	1	99.75	100	99.5
Bior2.6	3	100	100	100	2	98.25	96.5	100	1	99.5	100	99
Bior2.8	3	100	100	100	2	97.5	95.5	99.5	2	99.5	99.5	99.5
Bior3.1	3	100	100	100	1	95. 5	96	95	3	99.5	100	99
Bior3.3	3	100	100	100	3	95	93.5	96.5	2	99	99.5	98.5
Bior3.5	3	100	100	100	2	95.75	93	98.5	2	98.25	99.5	97
Bior3.7	3	100	100	100	2	95.75	93.5	98	2	99	99	99
Bior3.9	3	100	100	100	2	96	93.5	98.5	2	98.75	99.5	98
Bior4.4	3	100	100	100	3	97.5	95	100	2	98.5	99	98
Bior5.5	3	100	100	100	2	97.75	95.5	100	2	97.25	97	97.5
Bior6.8	3	100	100	100	3	96.5	93.5	99.5	2	98.5	99	98
Coif1	3	100	100	100	2	97.25	94.5	100	1	99.5	100	99
Coif2	3	100	100	100	3	97.5	95	100	2	98	100	96
Coif3	3	100	100	100	3	97	94	100	1	98	98	98
Coif4	3	100	100	100	3	97.75	95.5	100	2	98.5	99.5	97.5
Coif5	3	100	100	100	3	96.75	94	99.5	2	98.25	99	97.5
Sym2	3	100	100	100	3	97.75	96	99.5	1	99	100	98
Sym3	3	100	100	100	3	97.25	95	99.5	2	99.25	100	98.5
Sym4	3	100	100	100	3	96.75	95	98.5	2	98.75	100	97.5
Sym5	3	100	100	100	3	97.5	95.5	99.5	2	98.25	100	96.5
Sym6	3	100	100	100	3	97.25	95	99.5	3	98.75	100	97.5
Sym7	2	100	100	100	3	96.75	94.5	99	3	98	100	96
Sym8	3	100	100	100	3	97.5	95.5	99.5	2	98.75	100	97.5
Dmey	2	100	100	100	3	97.5	95	100	3	98.25	100	96.5

Table 8. The obtained results for seizure detection with the SVM classifier using the third method (extracting statistical features from the DWT coefficients of 37 different wavelets) with data samples D-E and ABCD-E.

	D-E				ABCD-E			
wavelet	**level**	**acc**	**sen**	**spe**	**level**	**acc**	**sens**	**spe**
Db1	3	96.5	100	93	1	97.5	95.5	98
Db2	3	96.25	99	93.5	2	97.4	95.5	97.87
Db3	3	96	99	93	3	97.2	94	98
Db4	3	96.5	99.5	93.5	2	97.2	94	98
Db5	3	96.5	99	94	3	97.3	95	97.87
Db6	3	96	98.5	93.5	3	97.4	93.5	98.37
Db7	3	96.75	99.5	94	3	97.1	95.5	97.5
Db8	3	96	99	93	3	97.1	95	97.62
Db9	3	96.5	99.5	93.5	3	96.9	94	97.62
Db10	3	96.25	99	93.5	3	97	94	97.75
Bior1.3	3	95.5	98	93	2	97.4	94	98.25
Bior1.5	3	96	98.5	93.5	1	97.2	92.5	98.97
Bior2.2	3	96.5	99	94	2	97.3	94	98.12
Bior2.4	3	96.5	99	94	3	97.3	93.5	98.25
Bior2.6	3	96.75	99.5	94	1	97.4	94.5	98.12
Bior2.8	3	96	98.5	93.5	1	97.3	94.5	98
Bior3.1	3	97	99	95	3	96.7	92	97.87
Bior3.3	3	95.5	98	93	3	96.3	91.5	97.5
Bior3.5	3	95.75	98	93.5	3	96.4	93.5	97.12
Bior3.7	3	96.25	99	93.5	2	96.2	92.5	97.12
Bior3.9	2	96.5	98.5	94.5	2	96.5	91	97.87
Bior4.4	3	96.5	99.5	93.5	2	97.4	94.5	98.12
Bior5.5	3	96.5	99	94	2	97.3	94.5	98
Bior6.8	3	96	99	93	3	97	94	97.75
Coif1	3	96	99	93	1	97.6	95.5	98.12
Coif2	3	97.25	100	94.5	3	97.6	94.5	98.37
Coif3	3	96	99.5	92.5	3	97.3	94.5	98
Coif4	3	96.25	99	93.5	3	97.3	94	98.12
Coif5	3	96.5	99.5	93.5	3	97.2	94.5	97.87
Sym2	3	96.25	99	93.5	2	97.4	95.5	97.87
Sym3	3	96	99	93	3	97.2	94	98
Sym4	3	95.75	98.5	93	3	96.7	93.5	97.5
Sym5	3	96	100	92	2	98	97.5	98.12
Sym6	3	96	99	93	3	97.1	93.5	98
Sym7	3	96.25	99.5	93	3	96.9	94	97.62
Sym8	3	96.25	99	93.5	3	96.7	92.5	97.75
Dmey	3	95.75	98	93.5	3	96.8	93	97.75

Table 9. The obtained results for Epilepsy detection with the SVM classifier using the DWT coefficients (db7 wavelet) after decomposing the EEG signal into five rhythms.

		A-C	A-D	B-C	B-D	AB-CD	AB-CDE
Delta rhythm	**Acc (%)**	92.5	93.5	83	95	92	94.3
	Sen (%)	91.5	90	66	96	90.5	92
	Spe (%)	93.5	97	100	94	93.5	97.75
Theta rhythm	**Acc (%)**	88	87.5	83.75	89.5	87.5	92.2
	Sen (%)	87.5	86	68	84.5	85.25	90.67
	Spe (%)	88.5	89	99.5	94.5	89.75	94.5
Alpha rhythm	**Acc (%)**	76.75	78	85.75	90.25	81.87	83.6
	Sen (%)	55	63	71.5	80.5	66.5	77.33
	Spe (%)	98.5	93	100	100	97.25	93
Beta rhythm	**Acc (%)**	78.75	80	82.5	87.5	80.75	83.5
	Sen (%)	58	60.5	65	75	62	74.33
	Spe (%)	99.5	99.5	100	100	99.5	97.25
Gamma rhythm	**Acc (%)**	78.25	80.25	83.75	84.75	80	79.9
	Sen (%)	58.5	61.5	67.5	69.5	61.25	78.67
	Spe (%)	98	99	100	100	98.75	81.75

Table 10. The obtained results for seizure detection with the k-NN classifier using the DWT coefficients (db7 wavelet) after decomposing the EEG signal into five rhythms.

		A-E	B-E	C-E	D-E	ABCD-E
Delta rhythm	**Acc (%)**	98.25	94.25	92.75	94.25	95.4
	Sen (%)	96.5	90.5	92.5	94.5	87
	Spe (%)	100	98	93	94	97.5
Theta rhythm	**Acc (%)**	99.5	98.25	95.75	94.75	96.9
	Sen (%)	99	96.5	95	93	89
	Spe (%)	100	100	96.5	96.5	98.87
Alpha rhythm	**Acc (%)**	99.5	88.5	97.25	97.75	94.9
	Sen (%)	99	87.5	98.5	97.5	84.5
	Spe (%)	100	89.5	96	98	97.5
Beta rhythm	**Acc (%)**	97	93.5	96.25	96.75	96.9
	Sen (%)	96.5	91	97.5	97	89.5
	Spe (%)	97.5	96	95	96.5	98.75
Gamma rhythm	**Acc (%)**	94.5	97	91.75	94.75	96.3
	Sen (%)	90.5	96.5	95.5	96	86
	Spe (%)	98.5	97.5	88	93.5	98.87

References

[1] iMotions (2019) *Electroencephalography The Complete Pocket Guide.*

[2] Johns Hopkins Medicine (n.d.) *Anatomy of the Brain.* Available at: https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-of-the-brain (Access date: 04-2020).

[3] Healthline (2015) *Cerebellum.* Available at: https://www.healthline.com/human-body-maps/cerebellum#1 (Access date: 04-2020).

[4] Joynt, R. (n.d.) *Brainstem.* Available at: https://www.britannica.com/science/brainstem (Access date: 04-2020).

[5] Muse (2018) *A Deep Dive into Brainwaves: Brainwave Frequencies Explained.* Available at: https://choosemuse.com/fr/blog/a-deep-dive-into-brainwaves-brainwave-frequencies-explained-2/ (Access date: 04-2020).

[6] Brainworks (n.d.) *What Are Brainwaves?* Available at: https://brainworksneurotherapy.com/what-are-brainwaves (Access date: 04-2020).

[7] Alzheimer's association (n.d.) *What Is Dementia?* Available at: https://www.alz.org/alzheimers-dementia/what-is-dementia (Access date: 04-2020).

[8] Mayo clinic (n.d.) *Dementia.* Available at: https://www.mayoclinic.org/diseases-conditions/dementia/symptoms-causes/syc-20352013 (Access date: 04-2020).

[9] Alzheimer's association (n.d.) *Inside the Brain, a tour of how the brain works.* Available at: https://www.alz.org/alzheimers-dementia/what-is-alzheimers/brain_tour_part_2 (Access date: 04-2020).

[10] MedicineNet (n.d.) *Epilepsy (Seizure Disorder).* Available at: https://www.medicinenet.com/seizure/article.htm#epilepsy_facts (Access date: 04-2020).

[11] WebMD (n.d.) *Schizophrenia and Your Brain.* Available at: https://www.webmd.com/schizophrenia/schizophrenia-and-your-brain#1 (Access date: 04-2020).

[12] Britton, J.W., Frey, L.C., Hopp, J.L. et al. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants.* Chicago: American Epilepsy Society. An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants, 2016.

[13] Wikipedia (n.d.) *Electroencephalography.* Available at: https://en.wikipedia.org/wiki/Electroencephalography#History (Access date: 04-2020).

[14] Stone, James L., Hughes, John R. (2013) *Early History of Electroencephalography and Establishment of the American Clinical Neurophysiology Society*, 'Journal of Clinical Neurophysiology', Vol 30 (1), pp. 28-44.

[15] Frye, C. (2016) *Describe the advantages and disadvantages of each the following experimental methods for cognitive neuroscience studies: PET, EEG, TMS.* Available at: http://charlesfrye.github.io/FoundationalNeuroscience//84/ (Access date: 04-2020).

[16] Farnsworth, B. (2019) *EEG vs. MRI vs. fMRI – What are the Differences?* Available at: https://imotions.com/blog/eeg-vs-mri-vs-fmri-differences/ (Access date: 04-2020).

[17] Mayo clinic (n.d.) *Epilepsy*. Available at: https://www.mayoclinic.org/diseases-conditions/epilepsy/diagnosis-treatment/drc-20350098 (Access date: 04-2020).

[18] Gandhi, T., Panigrahi, B. K., & Anand, S. *A comparative study of wavelet families for EEG signal classification*. 'Neurocomputing', Vol 74(17), pp. 3051-3057, 2011.

[19] Ullah, I., Hussain, M., Qazi, E. and Aboalsamh, H. *An Automated System for Epilepsy Detection using EEG Brain Signals based on Deep Learning Approach*. Expert Systems with Applications, Vol.107, pp.61–71, Octobre 2018.

Available at: http://dx.doi.org/10.1016/j.eswa.2018.04.021.

[20] Nicolaou, N. and Georgia, J. *Detection of epileptic electroencephalogram based on permutation entropy and support vector machines*. 'Expert Systems with Applications', Vol. 39(1), pp. 202-209, 2012.

Available at: https://doi.org/10.1016%2Fj.eswa.2011.07.008

[21] Parvez, M.Z. and Paul, M. Features Extraction and Classification for Ictal and Interictal EEG Signals using EMD and DCT', 15th International Conference on Computer and Information Technology (ICCIT), Chittagong, pp. 132-137, Dec., 2012.

Available at: http://dx.doi.org/10.1109/iccitechn.2012.6509719

[22] Bajaj, V. and Pachori, R.B. *Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of EEG signals*, 'Biomedical Engineering Letters', Vol.3(1), pp.17–21, 2013.

[23] Sharaf, Ahmed I., Mohamed, A. and El-Henawy , Ibrahim M. *An Automated Approach for Epilepsy Detection Based on Tunable Q-Wavelet and Firefly Feature Selection Algorithm*. International Journal of Biomedical Imaging. Vol.2018, pp.1–12, Sep.2018.

[24] Martis, R.J., Acharya, U.R., Tan, J.H. et al. *Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction*. International Journal of Neural Systems. Vol.23(5), pp. 1557–1565, 2013.

[25] Ahammad, N., Fathima, T., & Joseph, P. (2014). Detection of Epileptic Seizure Event and Onset Using EEG. BioMed Research International,1–7,2014.

Available at: http://dx.doi.org/10.1155/2014/450573

[26] Juárez-Guerra, E., Alarcon-Aquino, V. & Gómez-Gil, P. *"Epilepsy Seizure Detection in EEG Signals Using Wavelet Transforms and Neural Networks"*. New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, pp.261–269, 2014.

Available at: http://dx.doi.org/10.1007/978-3-319-06764-3_33.

[27] Lasefr, Z., Ayyalasomayajula, S. and Elleithy, K.). 'Epilepsy Seizure Detection Using EEG signals. At Department of Computer Science and Engineering', University of Bridgeport, Bridgeport, CT, USA, 2017.

[28] Peachap, A.B. and Tchiotsop D. (2019). *Epileptic seizures detection based on some new Laguerre polynomial wavelets, artificial neural networks and support vector machines.* Informatics in Medicine Unlocked, Vol. 16, pp. 100209, 2019.

[29] NEUROTECHEDU (n.d.). *Preprocessing*. Available at: http://learn.neurotechedu.com/preprocessing/ (Access date: 04-2020).

[30] Wikipedia, the free encyclopedia (2020) *Butterworth filter*. Available at: https://en.wikipedia.org/wiki/Butterworth_filter (Access date: 04-2020).

[31] Hayes, M. H. *Schaum's outlines: digital signal processing.* 2nd edn. New York : McGraw-Hil, pp. 367-368, 2012

[32] Hargrave, M. (2020) *Standard Deviation Definition*. Available at: https://www.investopedia.com/terms/s/standarddeviation.asp (Access date: 04-2020).

[33] Wikipedia, the free encyclopedia (2020) *Median.* Available at: https://en.wikipedia.org/wiki/Median (Access date: 04-2020).

[34] Corporate Finance Institute (n.d.) *Mode: A statistical measure of central tendency*. Available at: https://corporatefinanceinstitute.com/resources/knowledge/other/mode/ (Access date: 04-2020).

[35] Montis, K. and Pei, T. (2010) *First Quartile and Third Quartile*. Available at: http://web.mnstate.edu/peil/MDEV102/U4/S36/S363.html (Access date: 04-2020).

[36] khan Academy. (n.d.) *Interquartile range review*. Available at: https://www.khanacademy.org /math/statistics-probability/summarizing-quantitative-data/interquartile-range-iqr/a/interquartile-range-review (Access date: 04-2020).

[37] Salomon, D. *Data Compression: The Complete Reference*, 4th edn. California: Springer, pp. 298-299, 2007.

[38] Amara, G. (n.d.) *An Introduction to Wavelets*. Available at: https://www.eecis.udel.edu/~amer/CISC651/IEEEwavelet.pdf (Access date: 04-2020).

[39] Sripathi, D. *Efficient Implementations of Discrete Wavelet Transforms Using FPGAs*, Master thesis, Florida State University, Florida, 2003.

[40] Misiti, M., Misiti, Y. and Oppenheim, G. *Wavelet Toolbox™ 4 User's Guide*. MathWorks, Inc., 2009.

[41] Addison, P.S. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. London: Institute of Physics Publishing, 2002.

[42] Daamouche, A.*Optimized Wavelets for Classification, Application to Hyperspectral Image and ECG signal*, PhD thesis, M'Hamed Bougara University, Boumerdes, 2012.

[43] Al-Fahoum, A.S and Al-Fraihat, A.A.*Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains, Hindawi*, Vol 2014.

[44] Corinthios,M. (2009) *Signals, Systems, Transforms, and Digital Signal Processing with MATLAB*, USA: CRC Press, pp.1157-1158.

[45] Wikipedia, the free encyclopedia (2020) *Discrete Wavelet Transform.* Available at: https://en.wikipedia.org/wiki/Discrete_wavelet_transform (Access date: 04-2020).

[46] Wikipedia, the free encyclopedia (2020) *Haar Wavelet.* Available at: https://en.wikipedia.org/wiki/Haar_wavelet (Access date: 04-2020).

[47] Wang, S.H., Zhang, Y.D., Dong, Z. and Phillips, P.(2018) *Pathological Brain Detection: Brain Informatics and Health*, Springer Singapore, pp 85-91.

[48] Dogra, A., Goyal, B and Agrawal, S. *Performance Comparison of Different Wavelet Families based on Bone Vessel Fusion*. *Assian Journal of Pharmaceutics*, 10(4), pp.791-795, 2016.

[49] Wang, S.H., Zhan, T.M., Chen, Y., Zhan, Y., Yang, M., LU, H.M., Wang, H.N, Liu, B. AND Phillips, P. *Multiple sclerosis detection based on biorthogonal wavelet transform, RBF kernel principal component analysis, and logistic regression*, *IEEE Access*, vol.4, pp. 7567-7576, 2016.

[50] Merry, R.J.E. *Wavelet Theory and Applications*, A literature study, Eindhoven University of Technology, 2005.

[51] Expert System (2020) *What is Machine Learning? A definition.* Available at: https://expert system.com/machine-learning-definition/ (Access date: 05-2020).

[52] javaTpoint (n.d.) *Regression vs. Classification in Machine Learning.* Available at: https://w ww.javatpoint.com/regression-vs-classification-in-machine-learning (Access date: 04-2020).

[53] Marsland, S. *MACHINE LEARNING: An Algorithmic Perspective*. 2nd edn. CRC Press /Taylor & Francis Group, 2015.

[54] Alaliyat, S. *Video -based Fall Detection in Elderly's Houses*, Master thesis, Gjøvik University College, 2008.

Available at: https://www.researchgate.net/publication/267953942 (Access date: 04-2020).

[55] Zhang, Z. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med*, Vol. 4(11):218, 2016.

[56] Wikipedia, the free encyclopedia (2020) *k-nearest neighbors algorithm.* Available at: https: //en.wikipedia.org/wiki/K-nearest_neighbors_algorithm (Access date: 04-2020).

[57] Kantardzic, M. DATA MINING: Concepts, Models, Methods, and Algorithms, 3rd edn. New Jersey : IEEE Press, John Wiley & Sons, Inc. 2020.

[58] Kellerhals, S. (n.d) *INTRO TO MACHINE LEARNING IN R (K NEAREST NEIGHBOURS ALGORITHM) GETTING STARTED WITH MACHINE LEARNING IN R*. Available at: https://our codingclub.github.io/tutorials/machine-learning/ (Access date: 04-2020).

[59] Pupale, R. (2018) *Support Vector Machines(SVM) — An Overview*. Available at: https://tow ardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989 (Access date: 04-2020).

[60] Ertekin, S. (2009) *LEARNING IN EXTREME CONDITIONS: ONLINE AND ACTIVE LEARNING WITH MASSIVE, IMBALANCED AND NOISY DATA*, PhD thesis, The Pennsylvania State University.

[61] Bhuvaneswari, P., Satheesh Kumar, J. ' Support Vector Machine Technique for EEG Signals', *International Journal of Computer Applications*, Vol.63(13), 2013.

[62] Fletcher, T. *Support Vector Machines Explained*, London's Global University, 2008.

[63] Zhu, X. *Support Vector Machines*, Lecture notes, CS769 Advanced Natural Language Processing, 2010.

[64] Wikipedia, the free encyclopedia (2020) *Artificial neural network*, Available at: https://en. wikipedia.org/wiki/Artificial_neural_network (Access date: 04-2020).

[65] Mesa, J.A.S., Galan, C. and Hervas, C. 'The use of discriminant analysis and neural networks to forecast with a typical Mediterranean climate', *International Journal of Biometeorology* (49), pp.355-362; 2005. Available at: https://www.researchgate.net/publication/ 7947079 (Access date: 04-2020).

[66] Tiwari, S. (n.d.) *Activation functions in Neural Networks.* Available at: https://www.geeksforgeeks.org/activation-functions-neural-networks/ (Access date: 04-2020).

[67] Oppermann, A. (n.d.) *Activation Functions in Neural Networks.* Available at: https://www.deeplearning-academy.com/p/ai-wiki-activation-functions (Access date: 04-2020).

[68] ML Glossary (2017) *Activation Functions*. Available at: https://ml-cheatsheet.readthedocs.io/en/latest/activation_functions.html#relu (Access date: 04-2020).

[69] Luckert, M. and Schaefer-Kehnert, M.(2015) *Using Machine Learning Methods for Evaluating the Quality of Technical Documents*, Master thesis.

[70] C. M. Bishop (2006) *Pattern Recognition and Machine Learning*, Singapore: Springer.

[71] The Bonn dataset. Available at: http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&chang elang=3 (Access date: 10-2019).

[72] Siuly S, Li Y, Zhang Y. *EEG Signal Analysis and Classification*. Health Information Science. Springer International Publishing; 2016.

Available at: http://dx.doi.org/10.1007/978-3-319-47653-