DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

National Polytechnic school

Electronic Department

Master Thesis in Electronics

A comparison study on EEG signal classification using Component analysis (PCA, ICA) and Support Vector Machine (SVM)

Hadjer AZLI

Supervised by:

PhD Mourad ADNANE

Pr. Adel BELOUCHRANI

Presented in public on October 8, 2017

Jury members

President	Mr. M. S. AIT CHEIKH	Pr	ENP
Examiner	Mr. C. LARBES	Pr	ENP
Supervisors	Mr. M. ADNANE	PhD	ENP
	Mr. A. BELOUCHRANI	Pr	ENP

ENP 2017

DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

National Polytechnic school

Electronic Department

Master Thesis in Electronics

A comparison study on EEG signal classification using Component analysis (PCA, ICA) and Support Vector Machine (SVM)

Hadjer AZLI

Supervised by:

PhD Mourad ADNANE

Pr. Adel BELOUCHRANI

Presented in public on October 8, 2017

Jury members

President	Mr. M. S. AIT CHEIKH	Pr	ENP
Examiner	Mr. C. LARBES	Pr	ENP
Supervisors	Mr. M. ADNANE	PhD	ENP
	Mr. A. BELOUCHRANI	Pr	ENP

ENP 2017

Dedication

I dedicate this work to

My beloved father may Allah have mercy on him.

And my greatest mother.

Acknowledgement

Foremost, I thank Allah the all-powerful, for given us the courage, the patience, willingness and strength to deal with all the difficulties and obstacles that have risen through our path, throughout this work.

I would like to express my sincere gratitude to my advisors Mr. M. ADNANE and Mr. A. BELOUCHRANI for their continues support during my study. I would also like to thank the rest of my thesis committee for the honor of evaluating my modest work.

I would give my deep thanks to my family and my friends for their guide and mental support, and last but not least a *special* gratitude to *my greatest mother* and *my beloved father* may *Allah* have mercy on him.

ملخص

هذا العمل يهدف الى معالج إشارات بيانات التخطيط الدماغي EEG بإستعمال تقنية للتصنيف الاوتوماتيكي بطريقة آلة دعم الاشعة و ذلك لتصنيف حالة المريض : مرضي او عادي.

في هذا العمل بتطبيق تقنيات تحليل الاشارات و بدئنا بتطبيق تحويل المويجات (DWT) , مرفقة باستخراج المكونات الخصوصية (احصائيات) لتشكيل مصفوفة الميزات. تقنيتين لتخفيض البيانات (PCA, ICA) استعملا لتمثيل المعلومات في معلم جديد صغير الابعاد. و أخيرا تم تدريب خوارزمية SVM و إستخدامها على مجموعة من بيانات للاختبار و تصنيفها الى عادي او مرضي. مستوى اداء عمليات التصنيف لمختلف الطرق قدمت و قورنت و ذلك لتبيين و تحديد عملية التصنيف الممتازة .

كلمات مفتاحية: *اشارات التخطيط الدماغي (EEG), تحويل المويجات المنفردة (DWT), التحليل بالمكونات الرئيسية (PCA), التحليل بالمكونات المستقلة, آلة دعم الأشعة , الصرع.*

Résume

Cette étude a pour but de traiter le signal de l'électroencéphalogramme (EEG) en utilisant une méthode de classification automatisée avec la machine à vecteur de support (SVM), afin de catégoriser l'état du patient: épileptique ou non épileptique.

Nous avons utilisé des techniques d'analyse de signal, en commençant par l'application d' une décomposition en ondelettes discrète (DWT) sur le signal d'origine, suivie par l'extraction des paramètres caractéristiques (statistique), et la construction de la matrice caractéristique. Ensuite, deux techniques de réduction PCA et ICA ont été explorées pour représenter les données dans un nouvel espace distinct avec une dimension réduite.

Enfin, un algorithme SVM a été formé et utilisé sur un ensemble de données de test pour classifier : épileptique ou non. La performance du processus de classification utilisant ces différentes méthodes est présentée et comparée pour montrer l'excellent processus de classification.

Mot clés : *Electroencéphalogramme (EEG), Transformé en ondelettes discret (DWT), Analyse en composants principales (ACP), Analyse en composants indépendantes (ACI), Epilepsie*

Abstract

This study aims to analyze and process Electroencephalogram (EEG) signals using an automated classification method with Support vector machine (SVM), to categorize patient's seizure: epileptic or non-epileptic.

We employed a framework of signal analysis techniques, and we started by applying discrete wavelet decomposition (DWT) on the original signal, followed by extracting a set of statistical features and building the feature matrix. Next, a feature reduction PCA and ICA were explored to represent the data in a new distinct space with reduced dimension. Finally, an SVM

algorithm was trained and used upon a set of testing data to be classified: epileptic or not. The performance of classification process due to different methods is presented and compared to show the excellent classification process.

Keywords: *Electroencephalogram (EEG), Discrete Wavelet Transform (DWT), Independent Component Analysis (ICA), Principal Component Analysis (PCA), Support Vector Machine (SVM), Epileptic Seizure*.

.

Table of Contents

List of Figure

List of Table

List of Abbreviations

EEG	:	Electroencephalography
ECoG	:	electrocorticography
SEEG	:	stereoencephalography
fMRI	:	functional Magnetic Resonance Imaging
MEG	:	magnetoencephalography
PCA	:	Principal Component Analysis
ICA	:	Independent Component Analysis
SVM	:	Support Vector Machine
WT	:	Wavelet Transform
DWT	:	Discrete Wavelet Transform
CWT	:	Continuous Wavelet Transform
PC	:	Principle Component
LDA	:	Linear discriminant analysis
TNR	:	True Negative Ratio
TPR	:	True Positive Ratio

Introduction

Electroencephalography (EEG) is one of the most used techniques for the study of the electrical activity of the brain. It remains unavoidable for the diagnosis of diseases such as epilepsy. Epilepsy is the most common serious brain disorder in the world; it is a tendency to have recurrent seizures. A seizure occurs as a result of a sudden, usually brief, excessive electrical discharge in a group of brain cells. These discharges can occur in different parts of the brain.

There are many different methods of capturing information on brain structures and functions, and they are divided into invasive and non-invasive measurement. The common non-invasive techniques are functional Magnetic Resonance Imaging (fMRI), magnetoence-phalography (MEG) and electroencephalography (EEG). Among these methods, EEG is the most versatile and cost efficient solution. [**1**]

EEG is known as a non-invasive measurement technique because signals are recorded by placing the electrodes on the scalp over the skull. This technique has either inherently low temporal or low spatial resolution, and suffers from low signal-to-noise ratio and/or poor high frequency sensitivity. In contrast, the invasive technique of electrocorticography (ECoG) and (SEEG) provides brain signals that have an exceptionally high signal-to-noise ratio, less susceptibility to artifacts than EEG, and a high spatial and temporal resolution. ECoG involves measurement of electrical brain signals using electrodes that are implanted subdural on the surface of the brain [**2**].

Reliable analysis of electroencephalogram (EEG) signals is crucial that could lead the way to correct diagnostic and therapeutic methods for the treatment of patients with neurological abnormalities, and different seizures. The involving of electronic technology can help for patient diagnosis and make it easy to the neurologist to interpret. Using Support vector machine algorithm in structural manner can be helpful to classify the epileptic seizure from normal seizure, also providing some other signal analysis can increase the performance of the classifier.

CHAPTER 01

Chapter 1

Clinical Review

This chapter presents a general medical context of the brain, the different techniques for tracking brain signals and also the most abnormalities encountered.

1.1Anatomy of the brain

Brain is the most important functional organ, which controls and coordinated other muscles and nerves in our body, it composes of cerebrum, cerebellum, and brainstem. The cerebrum part is divided into two hemispheres known as left hemisphere and right hemisphere, each hemisphere is further divided into four lobes such as, Frontal, Temporal, Parietal and Occipital lobes see **Figure 1.**

- Frontal lobe is the largest lobe which is located behind the forehead.
- The left frontal lobe is responsible for speech and language. It concerned with planning, organizing, problem solving, memory, impulse control, decision making, selective attention and controlling behavior and emotions. If frontal lobe gets damage it may affect emotions, languages and memory.
- Temporal lobe is located on the sides of the brain under parietal and behind frontal lobe. This is responsible for sound and speech in various aspects of memory. It may create hearing, language and sensory problems during damage.
- Occipital lobe is placed at the lower back of the head which relates perception and process virtual information. It creates visual and perception defects affect getting injury on this part.
- Behind the frontal lobe, Parietal lobe is located with integrates sensory information from different parts of body. It may create the inability problem for recognize and locate parts of the body. [**3**]

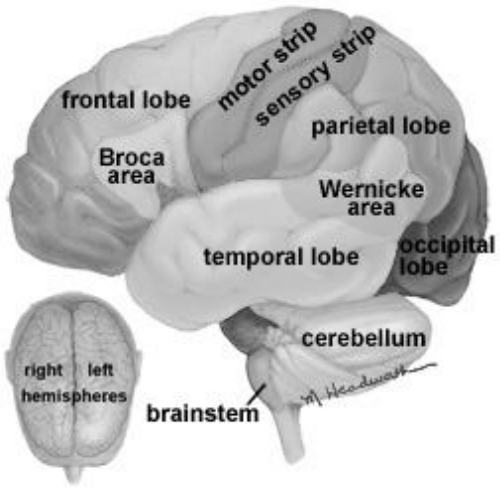

Figure 1: brain representation of the two hemispheres and of the different lobes

1.1.1 Cells of the brain

The surface of the cerebrum has a folded appearance called the cortex. The cortex contains about 70% of the 100 billion nerve cells. Beneath the cortex there are long connecting fibers between neurons, called axons, which make up the white matter.

Nerve cells: There are many sizes and shapes of neurons, but all consist of a cell body, dendrites and an axon. The neurons conveys information through electrical and chemical signals and transmit their energy, or "talk", to each other across a tiny gap called a synapse see **Figure 2.** The neuron has many arms called dendrites, which act like antennae picking up messages from other nerve cells. These messages are passed to the cell body, which determines if the message should be passed along. Important messages are passed to the end of the axon where sacs containing neurotransmitters open into the synapse. The neuro-transmitter molecules cross the synapse and fit into special receptors on the receiving nerve cell, which stimulates that cell to pass on the message.

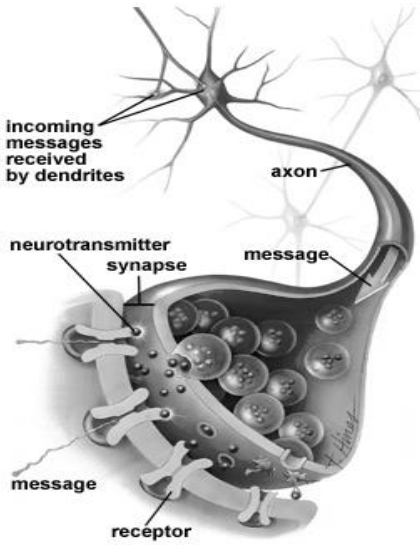

Figure 2: Nerve cell

1.1.2 Electromagnetic activity

The electrical potentials presented on the scalp are the consequences of the synaptic activity of neurons. The electrical activity of the brain collected on the surface of the scalp is due to the simultaneous activation of a very large number of neurons called Pyramidal cells.

For an electrical activity to be sufficiently broad to become visible on the scalp, hundreds of thousands of neurons must synchronize. The minimal cortical surface active for potential scalp appearance was estimated at $6\ cm^2$ [**4**].

1.1.3 Epileptic seizure

Epilepsy is a common brain disorder that, according to an estimate of the World Health Organization, affects almost 60 million people around the world. Some seizures are characterized by uncontrollable and rapid shaking of an individual's body due to an irregular or atypical electrical conductivity or connection in the brain. A seizure also can consist of any of the following: a blank stare, tremors or jerks, a convulsion with a total loss of consciousness, strange feelings and sensations, unusual tastes, lip-smacking and chewing, visual disturbances, aimless wandering, fiddling with clothes or objects. These behaviors and how they present all relate back to the area of the brain from which the seizure is originating.

Epilepsy is the most common form of convulsive or seizure disorder which can lead to dangerous and possibly life-threatening situations. The seizures are the result of a transient and unexpected electrical disturbance of the brain and excessive neuronal discharge that is evident in the electroencephalogram (EEG) signal representative of the electrical activity of the brain.

1.2 Measurement of brain activity

Today's technology provides many useful tools for studying the brain, some have their most important applications in medical diagnosis and some are used more for research.

There are two main groups of procedures: Structural and Functional analysis. Structural analysis is used to analyze the anatomy of the brain, in order to find structural deviations. Functional analysis tries to measure and locate brain activity, this is useful for investigating the functioning of special structures, and to diagnose epileptic seizures or diseases affecting brain activity [**5**] . In this thesis we are concerned about functional analysis especially the *electroencephalogram (EEG).*

1.2.1 Non-invasive methods

1.2.1.1 Electroencephalogram (EEG)

EEGs use small electrodes to detect voltage fluctuations at the scalp which are caused by the aggregate electrical activity of large numbers of neurons close beneath the scalp. When neurons fire together in synchrony, voltage fluctuations at the scalp have higher power; when they fire asynchronously, voltage fluctuations at the scalp have lower power. Oscillations can be filtered into different frequency bands, and the relative power of these different bands can be compared for different stimuli.

Different frequency bands are associated with different aspects of perceptual, motor, and cognitive function. The main frequencies of the human EEG waves are 5 groups: delta-δ ($0.5 - 4\ Hz$), theta-θ ($4 - 8\ Hz$), alpha-α ($8 - 13\ Hz$), beta-β ($13 - 30\ Hz$), and gamma-γ ($up\ to\ 30\ Hz$). **Table 1** Summarize these five waves and their characteristics:

Type	**Frequency (Hz)**	**Behavioral/Psychological State**	**Location**
Delta	0 – 4	Deep rest, Dreamless sleep	Frontally in adults, Posteriorly in children
Theta	4 – 8	Deeply relaxed	Thalamic region
Alpha	8 – 13	Day dream, calm	Posterior regions
Beta	13 – 30	Alert, active thinking, anxiety, panic attack, focus, concentration	Frontal and parietal
Gamma	30 - 100	Combination of two senses	Somatosensory cortex

Table 1: Brain wave frequencies with their characteristics [6]

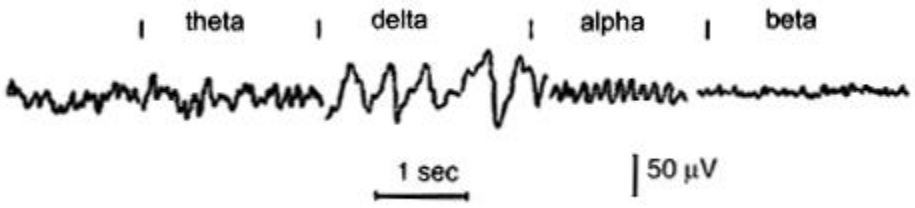

Figure 3: Theta, Delta, Alpha, Beta waveforms [7]

Variables used in the classification of EEG activity:

Frequency: Frequency refers to rhythmic repetitive activity (in Hz). The frequency of EEG activity can have different properties including:

- **Rhythmic**. EEG activity consisting in waves of approximately constant frequency.
- **Arrhythmic**. EEG activity in which no stable rhythms are present.
- **Dysrhythmic**. Rhythms and/or patterns of EEG activity that characteristically appear in patient groups or rarely or seen in healthy subjects.

Perturbations: The recorded activity which is not of cerebral origin is termed artifact and can be divided into physiologic (generated from the subject from sources other than the brain) and extra-physiologic artifacts arise from outside the body (equipment including the electrodes and the environment).

- **The artifacts:** Artifacts are non-cerebral electrophysiological activities recorded by the EEG (e.g. movement, sweating, ECG, eye movements).

    - **Eye movements:** The eyeball acts as a dipole with a positive pole oriented anteriorly (cornea) and a negative pole oriented posteriorly (retina). When the globe rotates about its axis, it generates a large amplitude alternate current field detectable by any of the electrodes positioned near the eye. A blink causes the positive pole (the cornea) to move closer to front polar FP1, FP2 electrodes, producing symmetric downward deflections.(**Figure 4**)

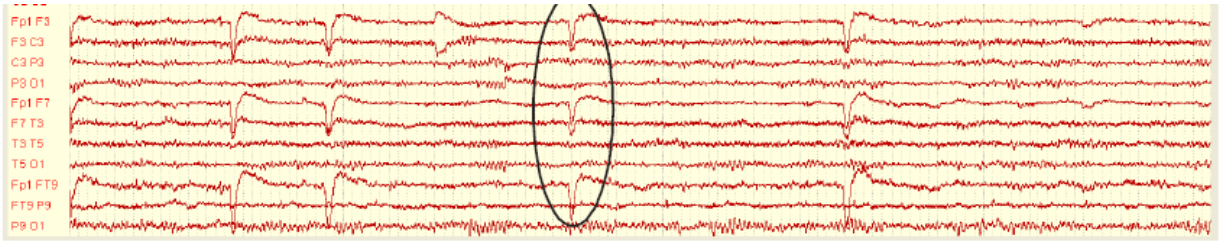

Figure 4: Eye blink artifacts

- **Skin artifacts**: A further difficulty arises due to properties of certain layers of the skin. A significant DC potential exists between the stratum corneum and the stratum granulosum and any local deformation of the skin will alter this potential. The only reliable way to eliminate the source of artifact is to create a low resistance pathway through the layers of skin by skin cleaning (alcohol swab). Also, sodium chloride (electrolyte) from sweating reacting with metals of the electrodes may produce a slow baseline drift.
- **Cardiac artifacts**: Each heartbeat causes a slight movement of the head so the electrodes located at proximity to an artery may be subject to further movement and generate an artifact [**4**]. See **Figure 5**

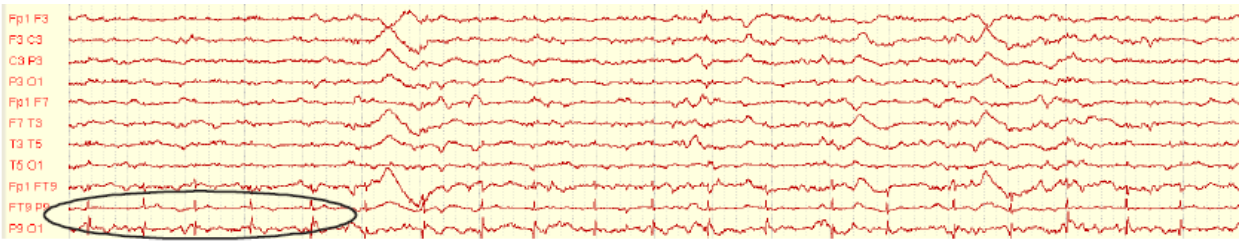

Figure 5: EEG data recording perturbed with cardiac activity

- **Electrodes:** Surface electrodes such as the ones used in EEG must create an interface between an ionic solution (the subject) and a metallic conductor (the electrode). This leads to a half-cell potential which can be quite large relative to the signal being recorded. To minimize this problem of polarization of the electrode, some electrodes are coated with silver chloride, but all are maintained away from the skin through an intermediate layer of conductive paste. Touching the electrodes during recording can produce artifacts. An electrode which is not contacting the skin very well acts like an antenna with resulting 60-cycle interference.

Electrode positioning (10/20 system)

The standardized placement of scalp electrodes for a classical EEG recording has become common since the adoption of the 10/20 system. The essence of this system is the distance in percentages of the 10/20 range between Nasion-Inion and fixed points. These points are marked as the Frontal pole (Fp), Central (C), Parietal (P), occipital (O), and Temporal (T). The midline electrodes are marked with a subscript z, which stands for zero. The odd numbers are used as subscript for points over the left hemisphere and even numbers over the right. [**7**]

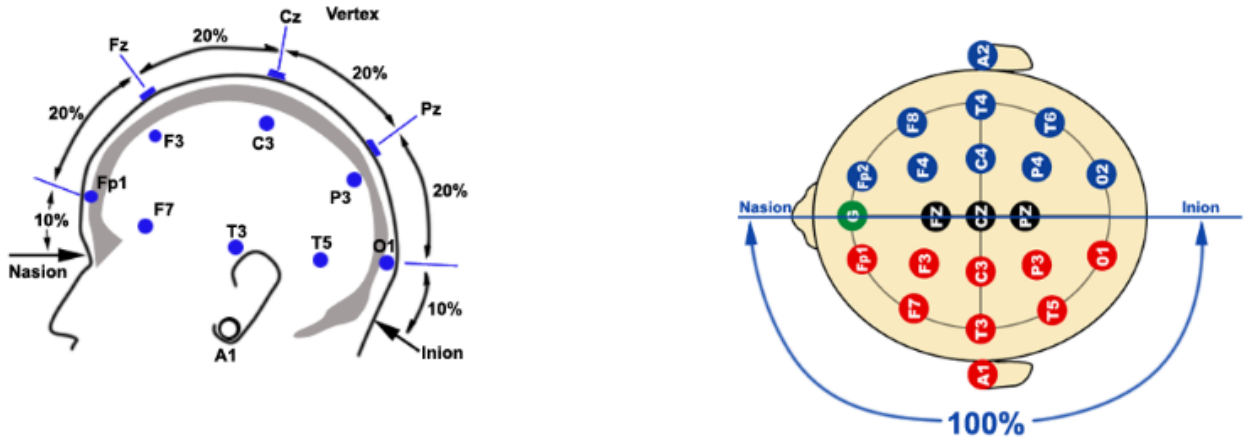

Figure 6: Electrode positioning

1.2.1.2 Functional magnetic resonance imaging (fMRI)

Functional MRI allows visualization of brain areas with increased blood flow during a functional task (or even during rest) .It takes advantage of the fact that oxygenated hemoglobin absorbs MRI signal, while deoxygenated hemoglobin does not. This change in blood flow is referred to as the blood-oxygenation level dependent (BOLD) or hemodynamic response. This increase in blood flow is taken as a proxy for activity because blood flow is known to be closely related to neural firing; active cells require more blood to support their activity.

fMRI has a great spatial and temporal resolution while it's weakness consist of that the images are acquired in a continuous stream during the behavior under study, so the behavior must be performed while the subject is immobilized inside the scanner. [**8**]

1.2.1.3 Magnetoencephalography (MEG)

MEG is similar to EEG except that it measures fluctuations in magnetic fields instead of electric fields .Instead of electrodes, MEG uses superconducting quantum interference device (SQUID) to detect these fluctuations.

MEG has very high temporal resolution. Otherwise it is not useful for measuring neural activity beneath the cortex, and its spatial resolution, while slightly better than EEG, is still relatively poor compared to fMRI.

1.2.2 Invasive methods

Invasive electroencephalography (iEEG) can be defined as electroencephalography (EEG) recording utilizing invasive methods or using invasive intracranial electrodes placed surgically.

The most invasive method is to record EEG by placing electrodes intracranially. Intracranial EEG recording can be performed either using electrodes placed directly on the exposed surface of the brain (subdural grid and strip electrodes) **Figure 8** , or by electrodes inserted into the brain parenchyma or within a lesion (depth electrodes) see **Figure 7** .

The recording from the cortical surface using subdural electrodes is referred to as electrocorticography (ECoG), whereas EEG recording using multiple depth electrodes is referred to as stereoencephalography (SEEG). [**9**]

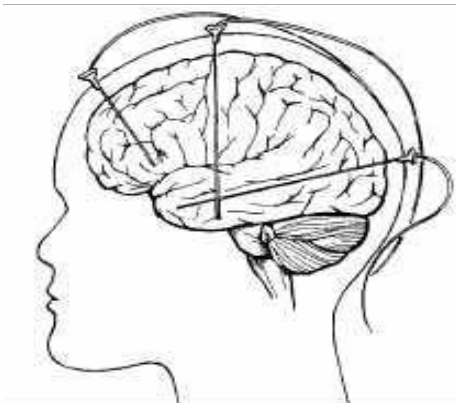

Figure 7: Cylindrical electrodes with sensors implanted in the brain (SEEG)

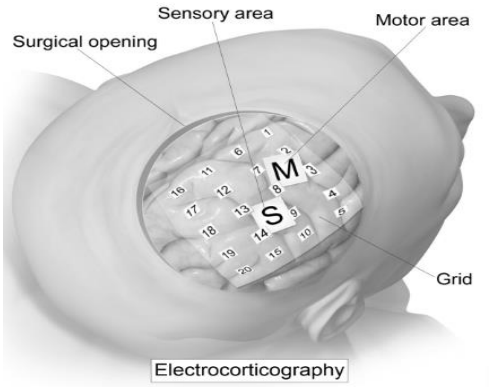

Figure 8: Intracranial electrode grid for electrocorticography (ECoG)

1.3 Conclusion

The detection of epileptic seizures by visual scanning of a patient's EEG data usually collected over a few days is a tedious and time-consuming process. In addition, it requires an expert to analyze the entire length of the EEG recordings, in order to detect epileptic activity. A reliable automatic classification and detection system would ensure an objective and facilitating treatment and significantly improve the diagnosis of epilepsy as well as long-term monitoring and treatment of patients. In the next Chapter we are going to state the different theoretical methods that involves in classification and system detection of epilepsy.

CHAPTER 02

Chapter 2

Theory and techniques

2.1 Signal analysis

EEG signal can be categorized to bands of different frequency ranges. Delta wave lies below the frequency of 4Hz. Theta lies in the range of 4Hz to 8Hz while Alpha wave lies between 8Hz to 13Hz. The range of Beta wave lies in 14Hz to 32Hz where beyond 32Hz lies the Gamma wave. These frequency bands each corresponds to different activities carried out by the subject. These different bands of frequencies each contain certain information of the brain activity. However, the information hides within the EEG signal is not directly analytical by the human eyes. However, information on neural connectivity may be revealed with the analysis of signal complexity on multiple scales. [**10**]

2.1.1 Wavelet transform

A signal is said to be stationary if it does not change much over time. Fourier transform can be applied to the stationary signals. However, like EEG, plenty of signals may contain non-stationary or transitory characteristics. By “windowing" the complex sinusoidal mother functions of the Fourier Transform, a time evolution of the frequencies can be obtained just sliding the windows throughout the signal. This procedure, called the Gabor Transform.

Gabor Transform gives an optimal time-frequency representation, but one critical limitation appears when windowing the data due to the Uncertainty Principle (see **Figure 9**). If the window is too narrow, the frequency resolution will be poor, and if the window is too wide, the time localization will be not so precise. Data involving slow processes will require wide windows and on the other hand, for data with fast transients (high frequency components) a narrow window will be more suitable. Then, owing to its fixed window size, Gabor Transform is not suitable for analyzing signals involving different range of frequencies.

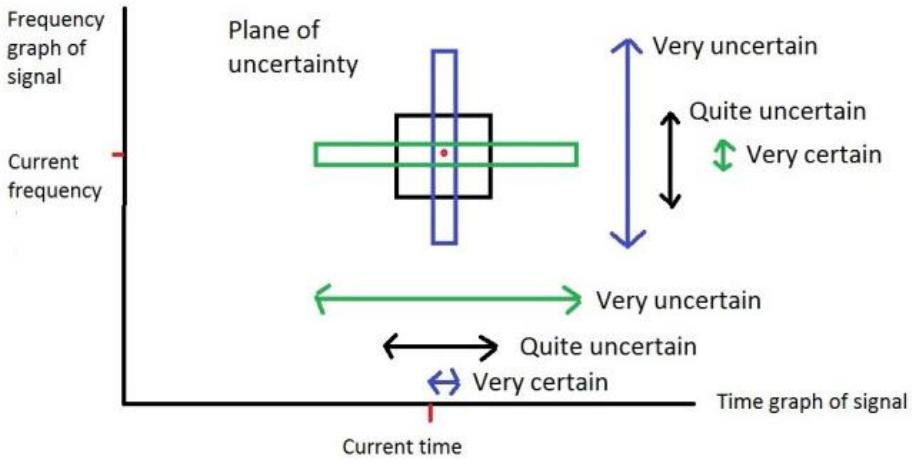

Figure 9: Plane of uncertainty

Grossmann and Morlet (1984) introduced the Wavelet Transform in order to overcome this problem. The main advantage of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones (see **Figure 10**), thus leading to an optimal time-frequency resolution in all the frequency ranges. [**11**]

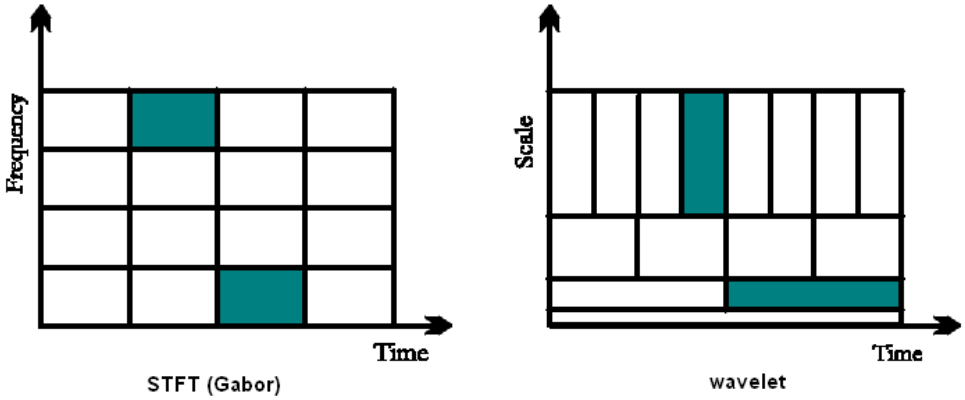

Figure 10: STFT and Wavelet windowing pattern

2.1.1.1 Continuous wavelet transform

The wavelet transform introduces an intriguing twist to the basic concept defined by Eq. **1** In wavelet analysis, a variety of different probing functions may be used, but the family always consists of enlarged or compressed versions of the basic function, as well as translations. This concept leads to the defining equation for the continuous wavelet transform (CWT):

$$W(a,b) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{|a|}}\psi * \left(\frac{t-b}{a}\right)dt \qquad 1$$

Where b acts to translate the function across $x(t)$ just as t does in the equations above, and the variable a acts to vary the time scale of the probing function. If a is greater than one, the wavelet function (ψ) , is stretched along the time axis, and if it is less than one (but still positive) it contacts the function. Negative values of a simply flip the probing function on the time axis. While the probing function ψ could be any of a number of different functions, it always takes on an oscillatory form, hence the term "wavelet." The $(*)$ indicates the operation of complex conjugation, and the normalizing factor $\frac{1}{\sqrt{a}}$ ensures that the energy is the same for all values of b as well, since translations do not alter wavelet energy. If $b = 0$, and $a = 1$, then the wavelet is in its natural form, which is termed the $Mother$ wavelet.

A mother wavelet is shown in **Figure 11** along with some of its family members produced by dilation and contraction. The wavelet shown is the popular *MORLET* wavelet.

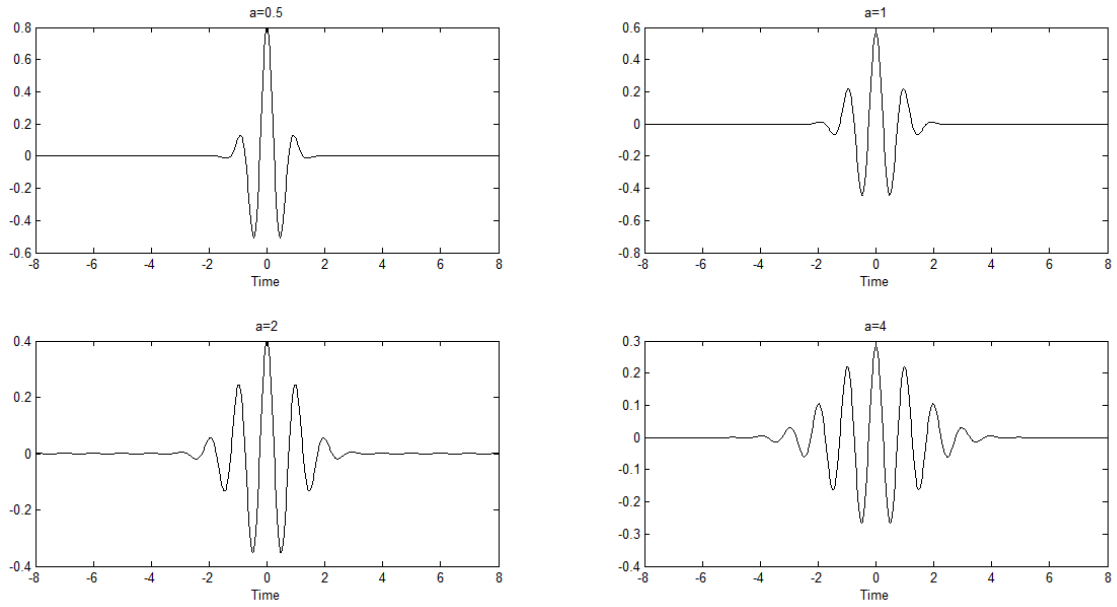

Figure 11: mother wavelet (a=1) with two dilatation (a=2, a=4) and one contraction (a=0.5)

The wavelet coefficients, $W(a,b)$ describe the correlation between the waveform and the wavelet at various translations and scales. In another way, the coefficients provide the amplitudes of a series of wavelets, over a range of scales and translations, that would need to be added together to reconstruct the original signal. To construct the original signal a double summation (or integration) is required see **Eq. 2** :

$$x(t) = \frac{1}{C} \int_{a=-\infty}^{\infty} \int_{b=-\infty}^{\infty} W(a,b)\psi_{a,b}(t)da\,db \qquad 2$$

Where :

$$C = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{|\omega|} d\omega \quad , \qquad 0 < C < \infty$$

The reconstruction of the original waveform is rarely performed using the CWT coefficients because of the redundancy in the transform. When recovery of the original waveform is desired, the more parsimonious discrete wavelet transform is used. [**12**]

2.1.1.2 Discrete wavelet transform (DWT)

The CWT has one serious problem: it is highly redundant. For recovery, all of the coefficients will be required and the computational effort could be excessive. In applications that require bilateral transformations, it would be preferred to produce a minimum number of coefficients required to recover accurately the original signal. The *discrete wavelet transform* (DWT) achieves this parsimony by restricting the variation in translation and scale, usually to powers of 2. The DWT is often introduced in terms of its recovery transform:

$$x(t) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d(k,l)2^{-\frac{k}{2}}\psi(2^{-k}t - l) \qquad 3$$

Where k is related to a as: $a = 2^k$; b is related to l as $b = 2^k l$; and $d(k,l)$ is a sampling of $W(a,b)$ at discrete points k and l .

Filter Bank

In the DWT, a new concept is introduced termed the *scaling function* represented in **Eq. 4** a function that facilitates computation of the DWT.

$$\phi(t) = \sum_{n=-\infty}^{\infty} \sqrt{2}c(n)\phi(2t - n) \qquad 4$$

The scaling function of the discrete wavelet transform can be represented as a tree of low and high pass filters (LP and HP), with each step transforming the low pass filter as shown in **Figure 12** . The original signal is successively decomposed into components of lower resolution,

while the high frequency components are not analyzed any further [**13**]. The Coefficients obtained after the HP filters are called detail coefficients while those after the LP filter are called the approximate coefficients.

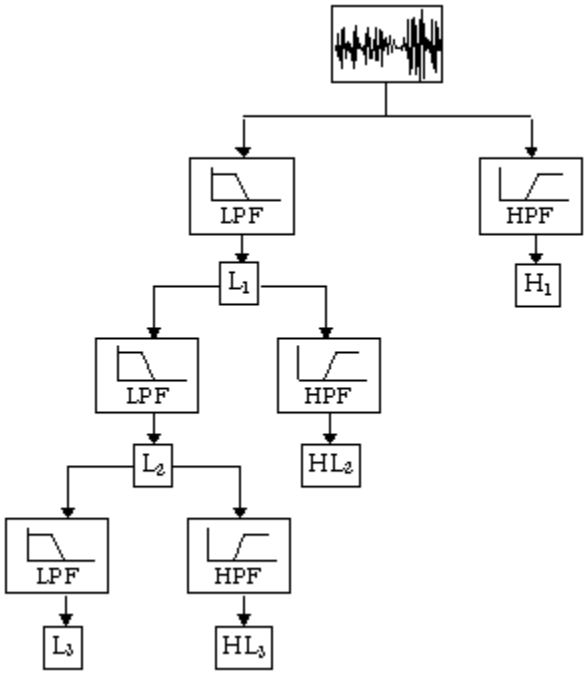

Figure 12: Filter Bank representation of the DWT Dilations

These groups of filters are used to divide up the signal into the spectral components called sub-band coding. The main parameter of the wavelet is to choose the number of levels of decomposition of the signal where these levels are based on the dominant frequency components of the signal. The theoretical expression of Maximum decomposition level is given as: [**14**]

$$DL_{max} = Log_2 N \qquad 5$$

2.2 Component analysis

2.2.1 Principal Component Analysis

Principal component analysis (PCA) is probably the most popular multivariate statistical technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in

maps. PCA seek to represent the data in a lower-dimensional space, this will reduce the degrees of freedom; reduce the space and time complexities.

Principal Component Analysis is often useful to measure data in terms of its PCs rather than on a normal x-y axis. They are the directions where there is the most variance, the directions where the data is most spread out. [**15**]

The data transformation used to produce the new set of variables is a linear function since linear transformations are easier to compute and their results are easier to interpret. A linear transformation can be represented mathematically as:

$$y_i(t) = \sum_{j=1}^{M} w_{ij} x_j(t) \quad , \qquad i = 1,2,\ldots,N \qquad \textbf{6}$$

Where W_{ij} is a constant coefficient that defines the transformation. Since this transformation is a series of equations, it can be equivalently expressed using the notation of linear algebra:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_M(t) \end{bmatrix} = \boldsymbol{W} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} \qquad \textbf{7}$$

As a linear transformation, this operation can be interpreted as a rotation and possibly scaling of the original data set in *M*-dimensional space. An example of how a rotation of a data set can produce a new data set with fewer major variables is shown in **Figure 13-A.**

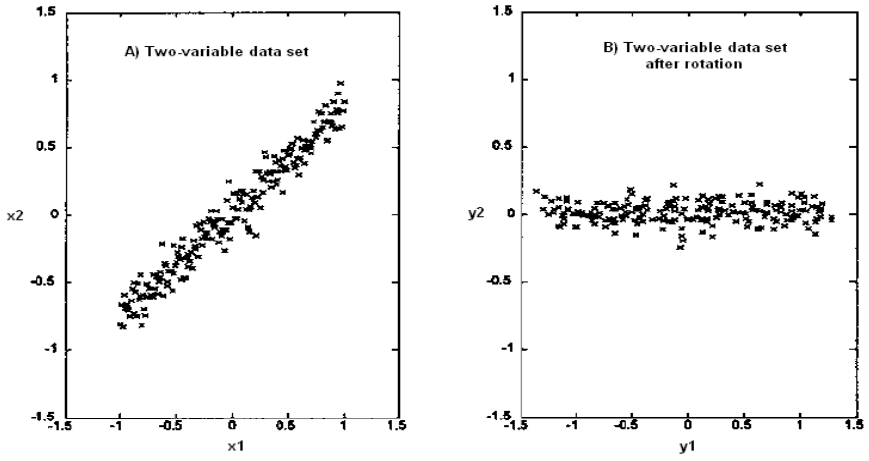

Figure 13: A data set consisting of two variables before (left graph) and after (right graph) linear rotation.

The original data set is shown as a plot of one variable against the other, a so-called *scatter plot*. In **Figure 13-B** the variance of variable y_1 and y_2 are greater than the variance of the first variable. This suggests that one variable,y_1 contains most of the information in the original two-variable set. [**12**]

Finding the components

In PCA, the goal is to find new set of variable with zero correlation, which means a diagonal covariance matrix. For that we can use technique from linear algebra given by the known formula:

$$S = VDV^T \quad \textbf{8}$$

Where $\boldsymbol{S}$ is the m-by-m covariance matrix, $\boldsymbol{D}$ is a diagonal matrix, and **V** is an orthonormal matrix that does the transformation . The rotation implied by $\boldsymbol{V}$ will produce a new covariance matrix $\boldsymbol{D}$, which has zero covariance. The diagonal elements of $\boldsymbol{D}$, are the variances of the new data, more generally known as the eigenvalues, of $\boldsymbol{S}$: $\lambda_1, \lambda_2, \dots, \lambda_N$. And the columns of V are the characteristic vectors, or eigenvectors.

We can get the eigenvalues and the eigenvectors for more variables by using singular value decomposition (SVD) **Eq. 9** or SVD economic given by **Eq. 10**

$$X = U\Sigma V^T \quad \textbf{9}$$

$$X = \widetilde{U}\,\widetilde{\Sigma}\,\widetilde{V}^T \ , \qquad \widetilde{\Sigma} : d \times d \quad \textbf{10}$$

In the case of data matrix $\boldsymbol{X_{n\times m}}$ where n is the number of observation and m is the number of features, $\boldsymbol{\Sigma}$ is a diagonal matrix which contains the square root of the eigenvalues; $\boldsymbol{V}$ is the eigenvectors matrix and called *Principal axes* or *Principal directions* of the data; the principle components F is given by:

$$F = XV = U\Sigma \quad \textbf{11}$$

The eigenvalues describe the variance accounted for by the associated principal components that are ordered by size and can be meaningful in identifying the number of principal components that are really significant. These principal components can then be used to reduce the data set as those contributing the least to the variance in the data can be eliminated.

2.2.2 Independent Component Analysis

Unlike principal component analysis, which is based on the assumptions of uncorr-elatedness and normality, ICA is rooted in the assumption of statistical independence. In another words Independent component analysis seeks to transform the original data set into number of independent variables and the motivation for this transformation is primarily to uncover more meaningful variables, not to reduce the dimensions of the data set.

For a random observed vector $X = \{X_1, X_2, \dots, X_m\}^T$ whose m elements are mixtures of m independent elements of a random vector $S = \{S_1, S_2, \dots, S_m\}^T$ given by:

$$X = AS \qquad \textbf{12}$$

Where A represents an $m \times m$ mixing matrix.

The goal of ICA is to find the unmixing matrix W that will give Y the best possible approximation of S:

$$Y = WX \cong S \qquad \textbf{13}$$

To estimate the mixing matrix, ICA needs only two assumptions: that the source variables S are truly independent; and that they are non-Gaussian except for one single source that can be Gaussian. Both conditions are usually met when the sources are real signals. A third restriction is that the mixing matrix must be square; in other words, the number of sources should equal the number of measured signals. This is not really a restriction since PCA can be always be applied to reduce the dimension of the data set, X to equal that of the source data set, S.

The Final restriction is termed in centering and whitening the data. Data that have been whitened are uncorrelated (as are the principal components), but in addition, all of the variables have variances of one. PCA can be used for both these operations.

The most significant computational difference between ICA and PCA is that PCA uses only second-order statistics (such as the variance which is a function of the data squared) while ICA uses higher-order statistics (such as functions of the data raised to the fourth power). Variables with a Gaussian distribution have zero statistical moments above second-order, but most *signals* do not have a Gaussian distribution and do have higher-order moments. These higher-order statistical properties are put to good use in ICA. [**16**,**12**]

2.3 Signal classification

Machine learning methods are helpful for interpreting high dimensional feature sets and analyze the characteristics of brain patterns. Support Vector Machine is one of the popular Machine Learning techniques for classifying the Electroencephalography (EEG) signals based on the neuronal activity of the brain.

2.3.1 Support Vector Machine (SVM)

SVM is a technique that applied in many applications like EEG signal classification, cancer identification, bioinformatics, seizure prediction, face recognition and speech disorder. It is one of the most recently developed classifiers in computational learning theory, because of their accuracy and their ability to deal with a large number of predictors.

Most of classifiers techniques separate classes using hyperplanes that split the classes using a flat plane, within the predictor space. Unfortunately, in many non-trivial problems, a perfect linear separator does not exist. SVMs extend the concept of hyperplane separation to data that cannot be separated linearly, by mapping the predictors onto a new, higher-dimensional space (called the feature space) in which they can be separated linearly see **Figure 14** .

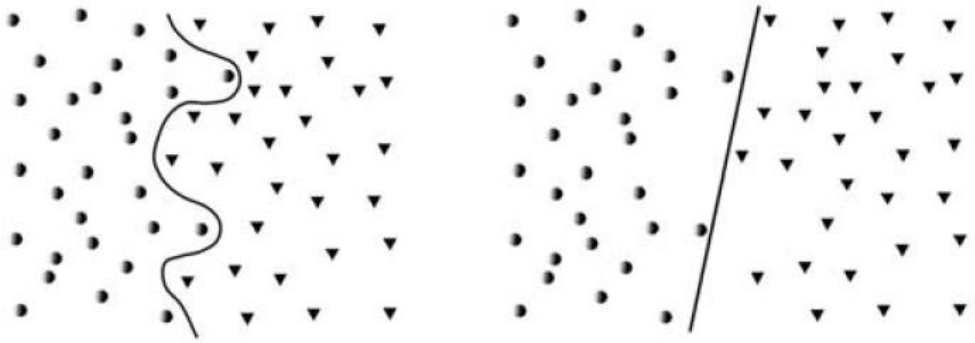

Figure 14: Diagrammatic representation of the hypothetical decision boundaries for two classes in the original (left) predictor space and transformed (right) feature space.

In this diagram the two classes can only be separated completely by a complex curve in the original space of the predictor. The best linear separator cannot completely separate the two classes. However, if the original predictor values can be projected into a more suitable “Feature space”, it is possible to separate completely the classes with a linear decision boundary. Consequently, the problem becomes one of finding the appropriate transformation. [**17**]

SVM can be classified into three types: linearly separable, Linearly Inseparable and Non linearly separable

2.3.1.1 Linearly Separable

Linearly Separable classification separates the high dimensional data into two groups, {+1,-1} without any overlapping or misclassification. The main objective of Support Vector Machine is maximizing the margin width in order to reduce the misclassification error.

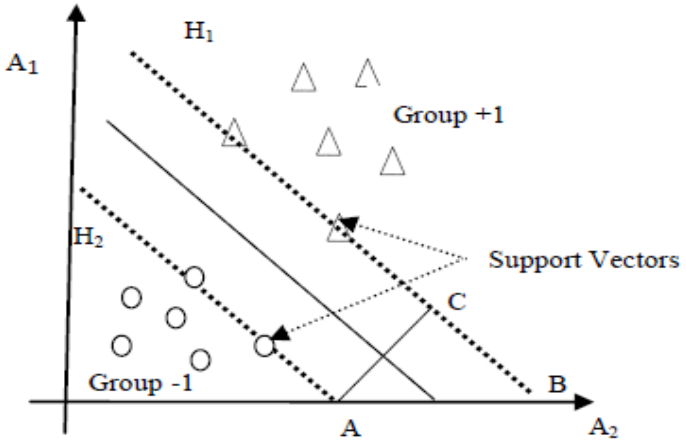

Figure 15: Linearly Separable

The linear classifier expression is given as:

$$h_{w,b}(x) = y(w^T x + b) \qquad 14$$

Where $y \in \{-1,1\}$, $w \in R^n$,$x \in R^n$ and b takes values $+1, 0, -1$ which shows how far hyperplanes away from the original line.

The maximum margin width is measured and it is equal to:

$$AC_{max} = \frac{2}{\sqrt{(w_1^2 + w_2^2)}} = \frac{2}{\|w\|} \qquad 15$$

Where w_1 ,w_2 are positions of the hyperplane H1 and H2 respectively, and x_1, x_2 are data points. [**6**]

2.3.1.2 Linearly Inseparable

In this case it is impossible to construct a linear hyperplane without error for binary classification data. Linearly inseparable classification can produce solutions for high dimensional datasets with overlapped or misclassified or erroneous data.

2.3.1.3 Non Linearly Separable

The major advantage of $SVMs$ is that they allow a non-linear enlargening of the feature space, while still retaining a significant computational efficiency, using a process known as the "kernel trick". In particular a set of p features $x_1, \dots, x_p$ can be transformed into a set of $2p$ features $x_1, x_1^2, \dots, x_p, x_p^2$. this allows us to apply a linear technique to a set of non-linear features. [**18**]

The higher dimension space is called feature mapping and its mapping function is denoted as $\phi(x_i)$. Kernel functions are used to find the value of mapping function ϕ .

$$x_i^T x_j = k(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad \textbf{16}$$

$k(x_i, x_j)$ Is called the Kernel functions which is based on the inner product of two variants x_i and x_j . [**6**]

Popular Kernel function

- Radial Kernel Function (RBF)

$$K(x_i, x_j) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \textbf{17}$$

- Linear Kernel Function

$$K(x_i, x_j) = {x_i}^T x_j \quad \textbf{18}$$

- Polynomial Kernel Function

$$K(x_i, x_j) = \left[\left({x_i}^T x_j\right) + 1\right]^d \quad \textbf{19}$$

- Gaussian Function

$$K(x_i, x_j) = \exp\left(-\frac{\left[\|x_i - x_j\|^2\right]}{2\,\sigma^2}\right) \quad \textbf{20}$$

2.4 Conclusion

In this chapter we sited different methods and technics of signal analysis that we would use in particular framework to classify the EEG data signal as epileptic and non-epileptic. And to achieve this structure a DWT, feature extraction, reduction technics and SVM were done on a set of data (S and O) of healthy and unhealthy patients.

So in chapter three we are going to apply those theoretical techniques and discuss their advantages on the way the classification was optimal or not.

CHAPTER 3

Chapter 3

Simulation and Results

In this chapter we are going to identifying the most suitable combination of dimensionality reduction technique paired with SVM that gave the highest sensitivity and specificity in classifying epileptic and non-epileptic data.

3.1 Methodology

3.1.1 Data collection

The data used in this test are from the publicly available data described by Andrzejak et al. (2001). The dataset consists of five subsets (denoted as S, F, N, O and Z) recorded with the same 128-channel amplifier system and 12-bit analog-to-digital convertor. Each of the subsets includes 100 segments with a sampling frequency of 173.61 Hz and duration of 23.6 s, i.e., 4096 sample points, the corresponding Nyquist frequency bandwidth is 86.8 Hz. EEG samples in Datasets O and Z are obtained from five healthy volunteers through external surface electrodes for open and closed eye conditions, while Datasets S, F and N consists of EEG segments recorded from epileptic patients using intracranial electrodes to monitor interictal and ictal epileptic activity. The datasets F and N are acquired during seizure free intervals, while the dataset S only contains the seizure activity. All five Datasets S, F, N, O and Z are tested and categorized into two separate groups using our proposed method. The epileptic seizure (S) class is composed of subset S, and the non-seizure (FNOZ) class includes Subsets F, N, O and Z, respectively [**19**]. In this study we used two subset (S and O), **S** for the epileptic class and **O** for non-epileptic.

3.1.2 Framework

Our aim in this work is to perform a comparison between PCA, ICA by using Support vector machine. To achieve that the algorithm steps are:

- Each segment of channel was normalized and decomposed on wavelets using discrete wavelet decomposition DWT.
- Extract features from each channel and compose feature matrix
- Use Principal Component analysis and Independent Component analysis to reduce the number of features of the feature matrix to decrease runtime..
- The reduced feature matrix is used as input to train the Support vector machine function.
- Perform a comparison between SVM trained to separate epileptic and non-epileptic signals on a test data set. The training matrix model used are:
    - PCA training matrix.
    - ICA training matrix.
    - The original Feature matrix without reduction.

To compare the performance of each model, a sensitivity and specifity and accuracy are used as performance measurement. Sensitivity and specifity are calculated from the result of data test classification.

Sensitivity value or the true positive ratio defined by dividing the true positive results numbers to total diagnosis numbers that are stated by the expert neurologists.

$$Sensitivity = TPR = \frac{TP}{TP + FN} \times 100\% \quad (21)$$

Specifity value or true negative ratio is the ratio of true negative results numbers to total diagnosis numbers stated by expert.

$$Specifity = TNR = \frac{TN}{TN + FP} \times 100\% \quad (22)$$

The accuracy is measured by the percentage of correct classification:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (23)$$

- TP = Number of correctly classified positive samples
- TN = Correctly classified negative samples while
- FP = Negative sample being classified as positive
- FP = Positive sample classified as negative.

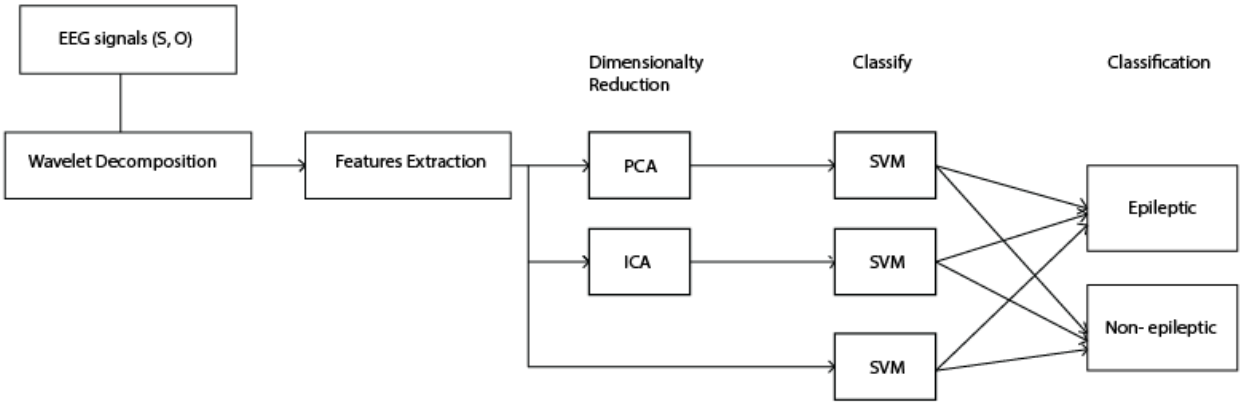

Figure 16: diagram of classification process

3.2 Results

3.2.1 Discrete wavelet Decomposition:

We performed wavelet decomposition on all channels of the epileptic and non-epileptic patient.

The signals are divided into high pass (Hi_D) and low pass (Lo_D) spectral characteristics where the high pass filter is like applying a wavelet to the original signal and the low pass filter is like applying a scaling or smoothing function.

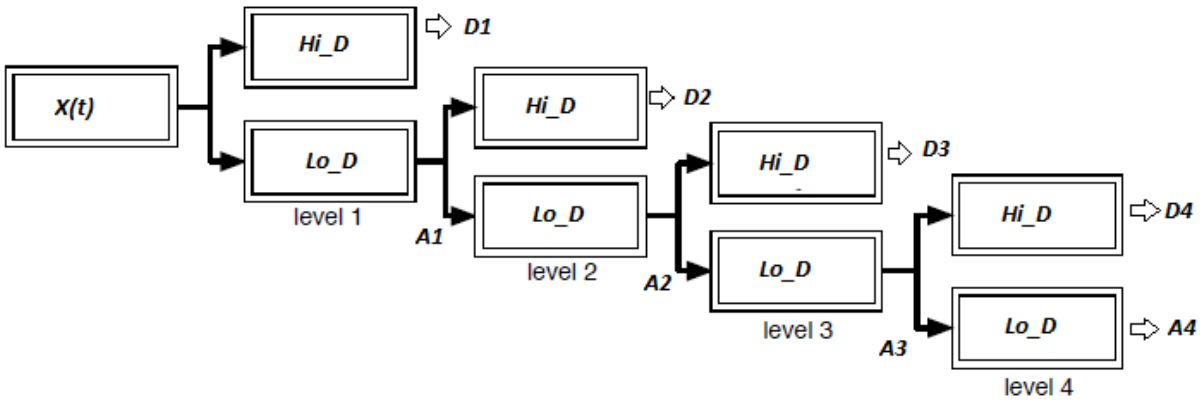

Figure 17: Sub-Band Decomposition

The number of levels was chosen based on the dominant frequency components of the signal. The sampling frequency was $173.61\ Hz$ and the maximum frequency existing in signals are less than 86 Hz (Nequist theory), as for that we decomposed the signal on five levels. Table 2 shows the decomposed signal and its frequency range.

Decomposed signal	Frequency range (Hz)
D1	43.4-86.8
D2	21.7-43.4
D3	10.8-21.7
D4	5.4-10.8
D5	2.7-5.4
A5	0-2.7

Table 2: Frequencies corresponding to different levels of decomposition

The approximation and detail coefficient are reconstructed from the Daubechies 6 (DB6) wavelet filter. Matlab functions used are: **wavedec()** for Wavelet decomposition, **appcoef()** , **detcoef()** to extract approximation and detailed coefficient, and **wrcoef()** to reconstruct the signals in time domain.

Figure 19 and Figure 21 shows wavelet decomposition of the epileptic data from one channel data set, Figure 18 and Figure 20 shows the original data.

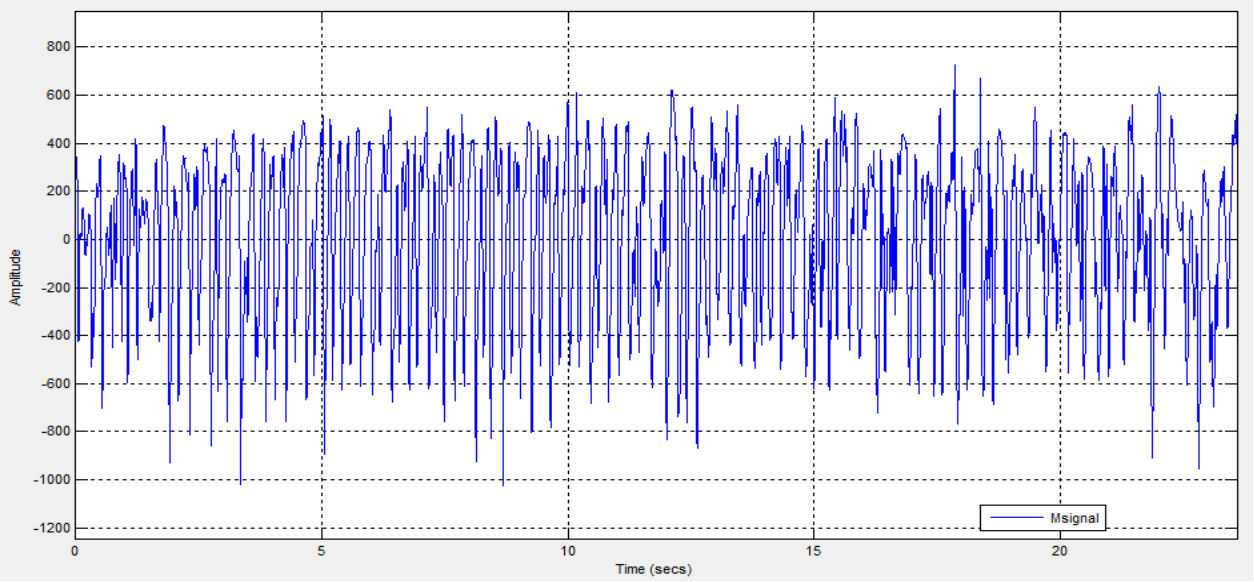

Figure 18: epileptic data from subset (S)

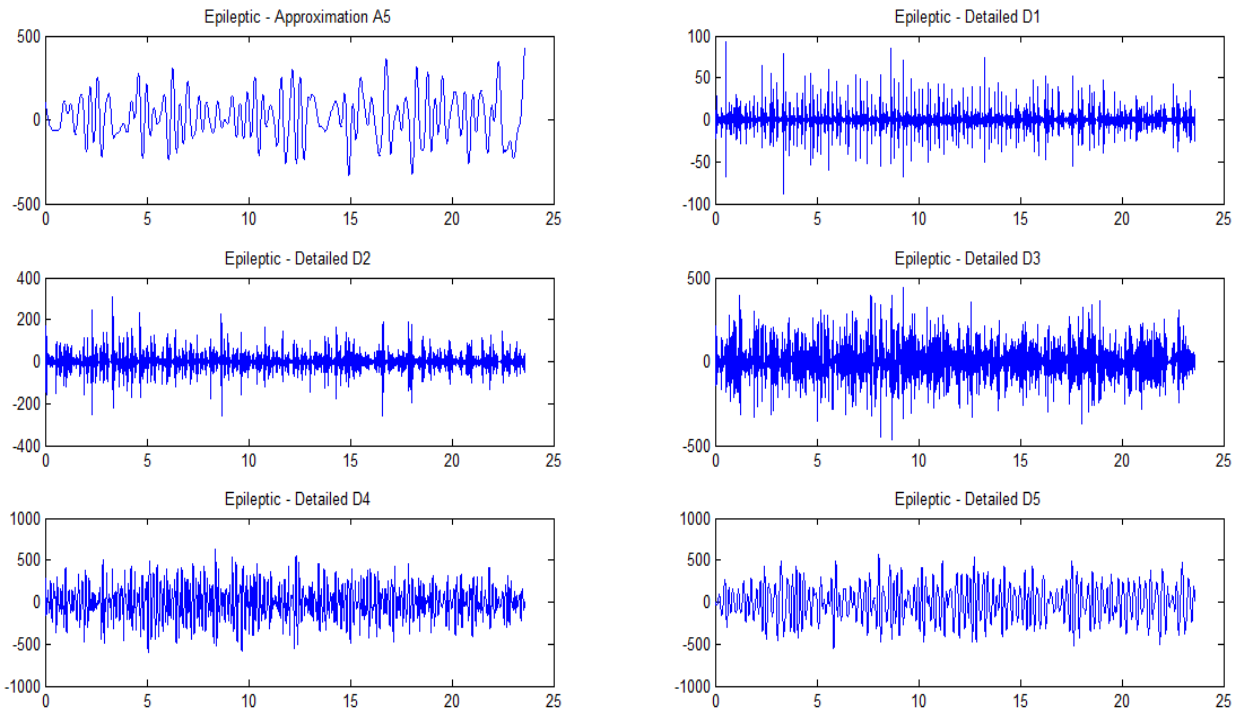

Figure 19: wavelet decomposition of epileptic data

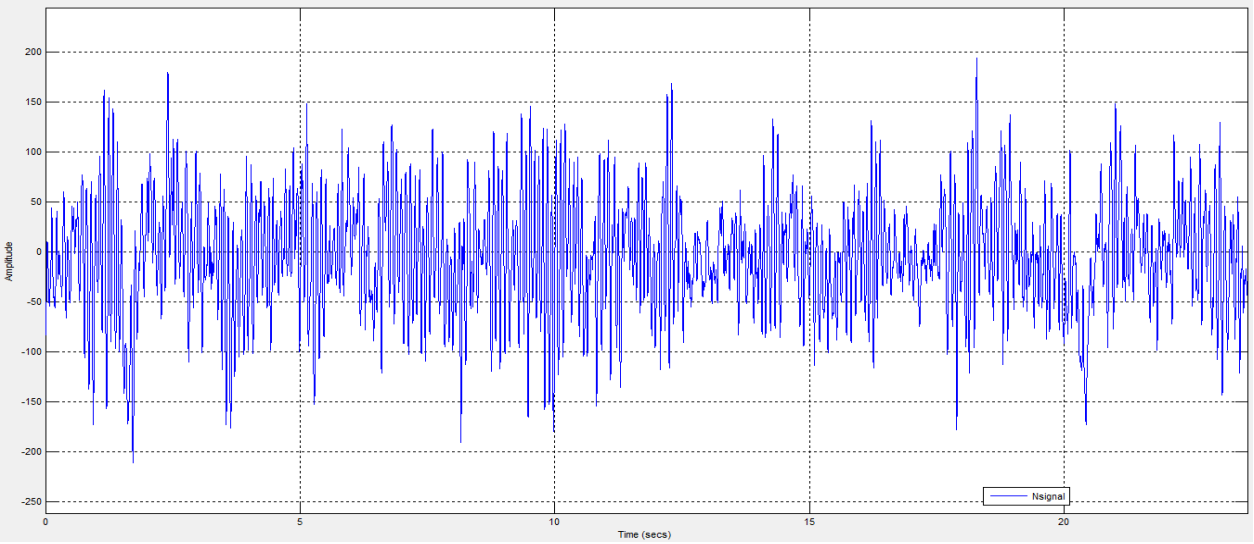

Figure 20: non-epileptic data from subset (O)

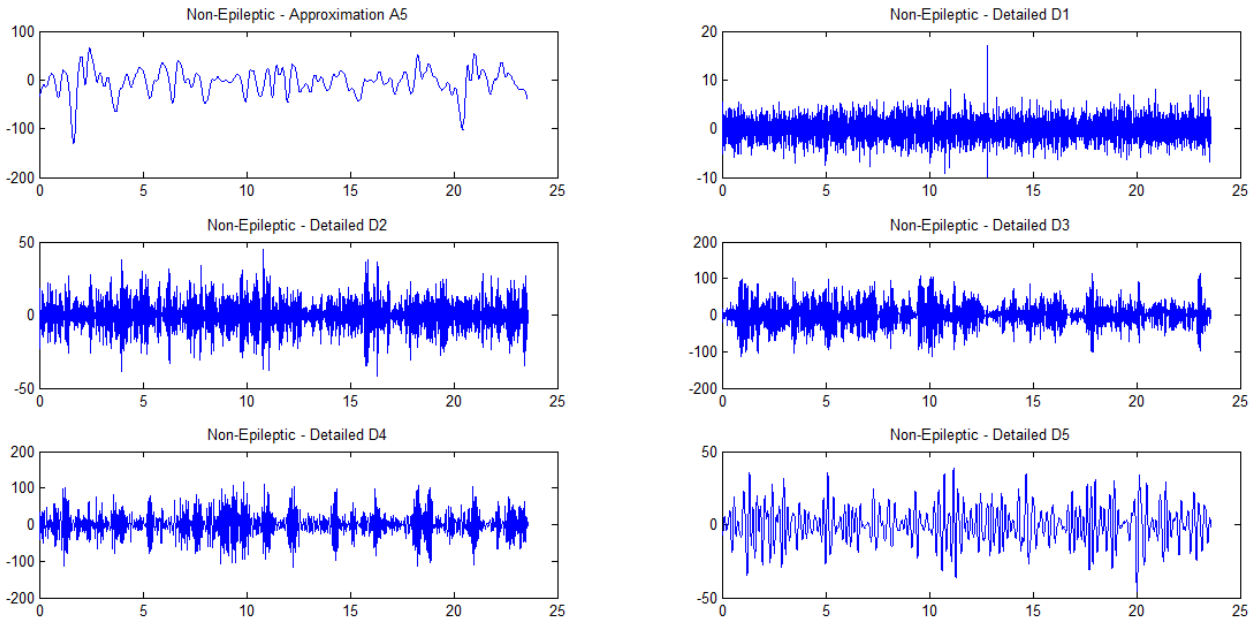

Figure 21: wavelet decomposition of non-epileptic data

3.2.2 Feature Matrix

The frequency range of interest that contains the four signal information: $\delta, \alpha, \beta, and\ \theta$ is from $0 - 30Hz$, thus the feature matrix represent the signal using statistics over the set of wavelet coefficients D2-D5 and A5.

The statistical features were used to represent the time frequency distribution of the EGG signals. Those features were chosen according to the method explained in the paper written by Subsei and Gursoy [**19**] and Tzanetakis et al. (2001) [**20**] :

- The mean of the absolute value of the coefficients in each sub-band. These features provide information about the frequency distribution of the signal.
- Average power of the wavelet coefficients in each sub-band.
- The standard deviation of the coefficients in each sub-band. These features provide information about the amount of change of the frequency distribution
- Ratios of the mean values between adjacent sub-bands. These features also provide information about the frequency distribution.

Each data segment of 4096 samples were divided into sub segment of 128 samples and an overlap of 64 samples between adjacent sub-segment.in this case each data segment results with a feature matrix of 63 rows and 19 columns. The rows present the time points (one row for each 128 samples) and the columns present the 19 statistical features.

Matlab function "**Feature_ex"** in Appendix A was used to calculate and generate feature matrix of the selected data signal.

3.2.3 Feature reduction

The feature matrix created contains 63 time point and 19 statistical features. The PCA function were used generate new principle component that are orthogonal to each other and has maximum variance of the data. Matlab function $PCA()$ was used to generate the representation of the data on the principle component space, also generate the percentage of the total variance explained by each principle component.

Figure 22 and Figure 23 show the scree plot of the variance explained for each PC of the two types of patients. After plotting all the variances from 100 segments, we observe a break in slope 5 and the total variance of the first five components is 80%, thus we suggest the number of PC of the data set to be five.

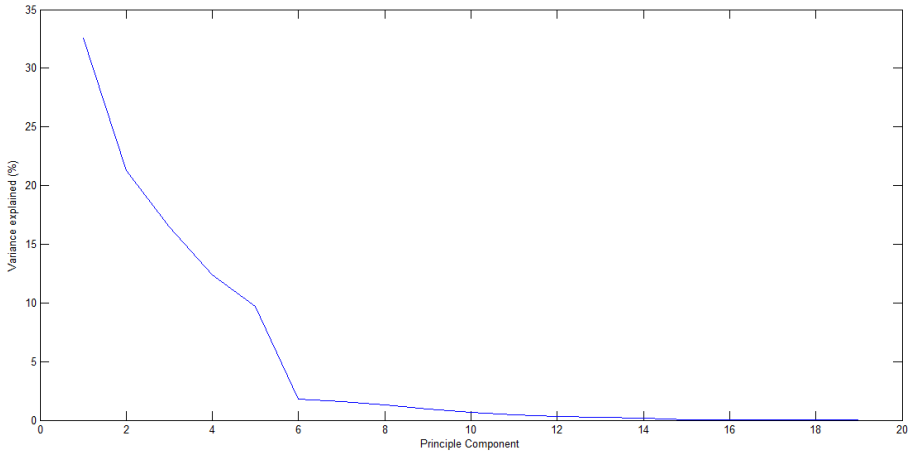

Figure 22: variance explained of epileptic data

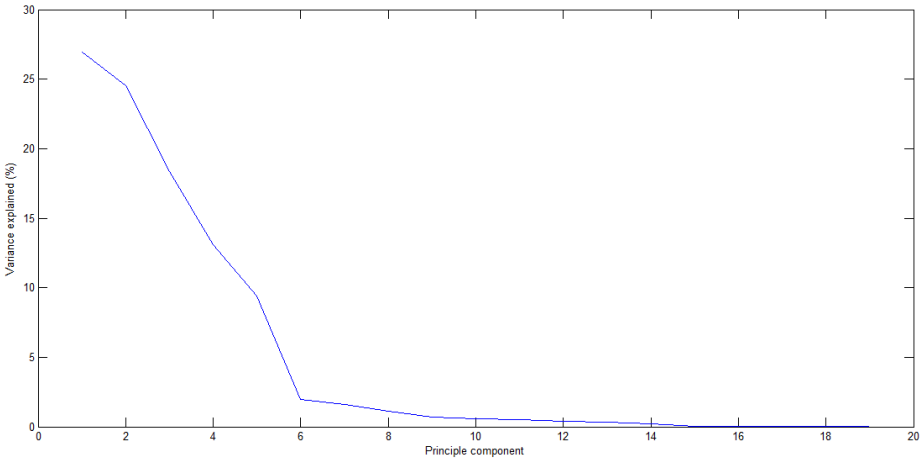

Figure 23: Variance explained of non-epileptic data

3.2.4 SVM training

SVM is the most popular machines learning tool that can classify data separated by non-linear and linear boundaries. In order to solve non-linear problems, SVMs use a kernel function which allows better fitting of the hyperplane to more general datasets.

A training matrix was fed into the SVM by using Matlab function SVMtrain() , after that a test dataset was used to test the SVM and that with the function SVMclassify(). Cross-validation techniques were used to get optimal classification of data.

The testing matrix was constructed from the 40 channels and the training matrix from 60 channels .The table below resume the channels of interest in training and testing matrix.

Class	**Training matrix**	**Test data set**
Epileptic dataset (S)	60 Channels	40 Channels
Non- epileptic dataset (O)	60 Channels	40 Channels

Table 3: Class distribution of the channels in the training and data sets

SVM classification results:

Figure 24 and **Figure 25** Shows classification of one channel (63 time point), the time points with negative unite ($y = -1$) are classified as epileptic, the positive unite ($y = 1$) are the non-epileptic class.

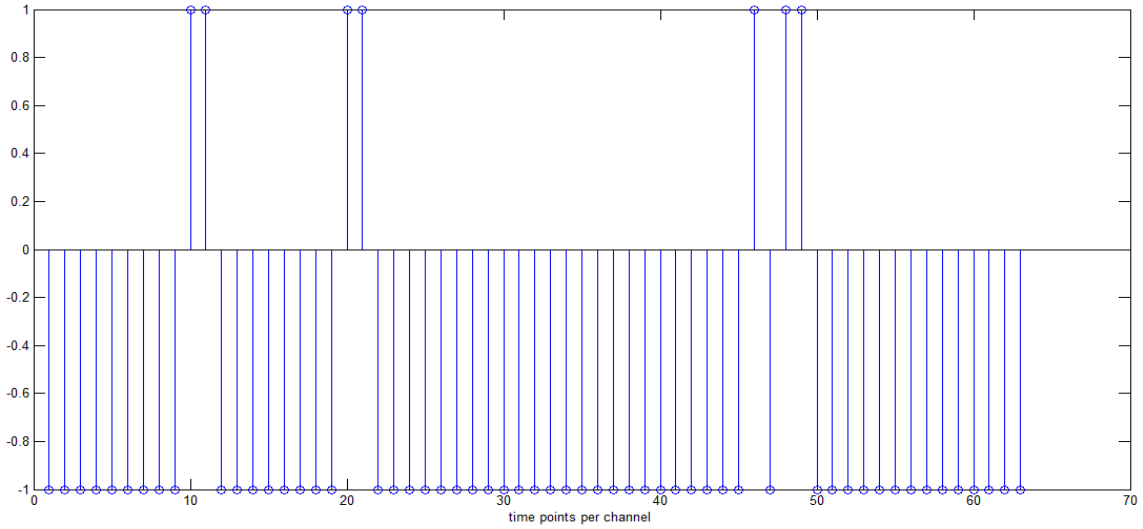

Figure 24: classification of epileptic channel

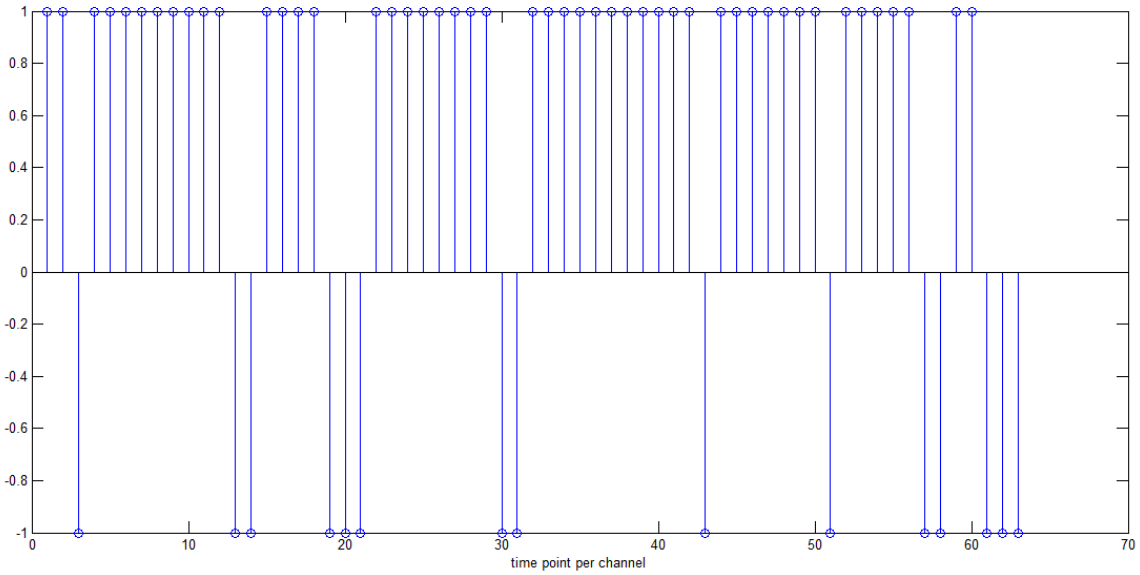

Figure 25: classification of Non-epileptic channel

Testing using model1 (original data):

The training matrix contains 60 channels of epileptic data and 60 of non-epileptic, data that resumes a total of 7560 time point. Matrix dimensionality was not reduced and it was fed directly to SVM. The test data to classify has 40 first channel epileptic and the last 40 channel non-epileptic. By knowing the true diagnosis of the test data we can calculate the performance of the learning machine according to the classification faults. The testing matrix must have the same dimensionality of the training matrix. **Figure 26** shows the result of classification using original features without reduction.

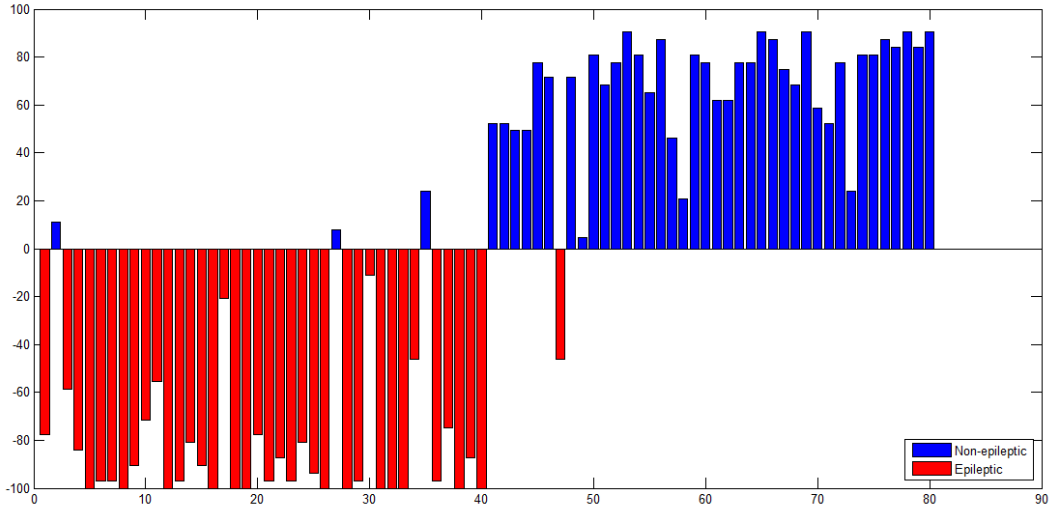

Figure 26: Channels classified without reduction techniques

The sensitivity and specificity of this model was calculated see Table 4 .

Testing Model 2(PCA):

The features were reduced to be five features. The new data in this five principle component are uncorrelated and have maximum variance. The classification results are in Figure 27

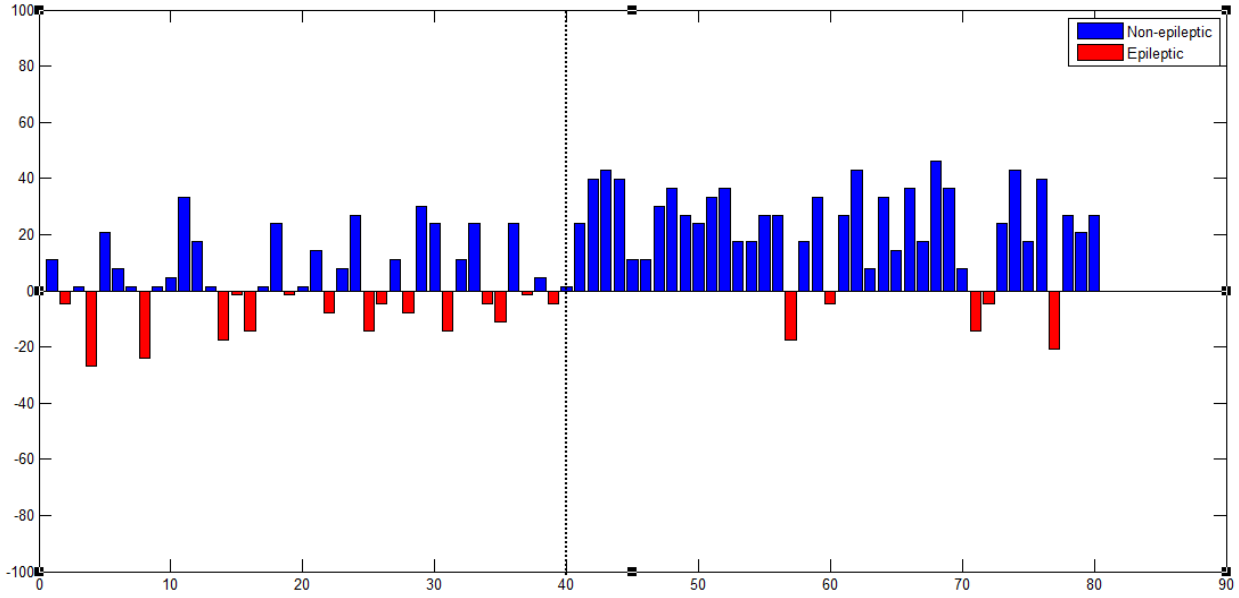

Figure 27: classification of 80 epileptic and non-epileptic Channels with PCA feature reduction

Testing Model 3 (ICA):

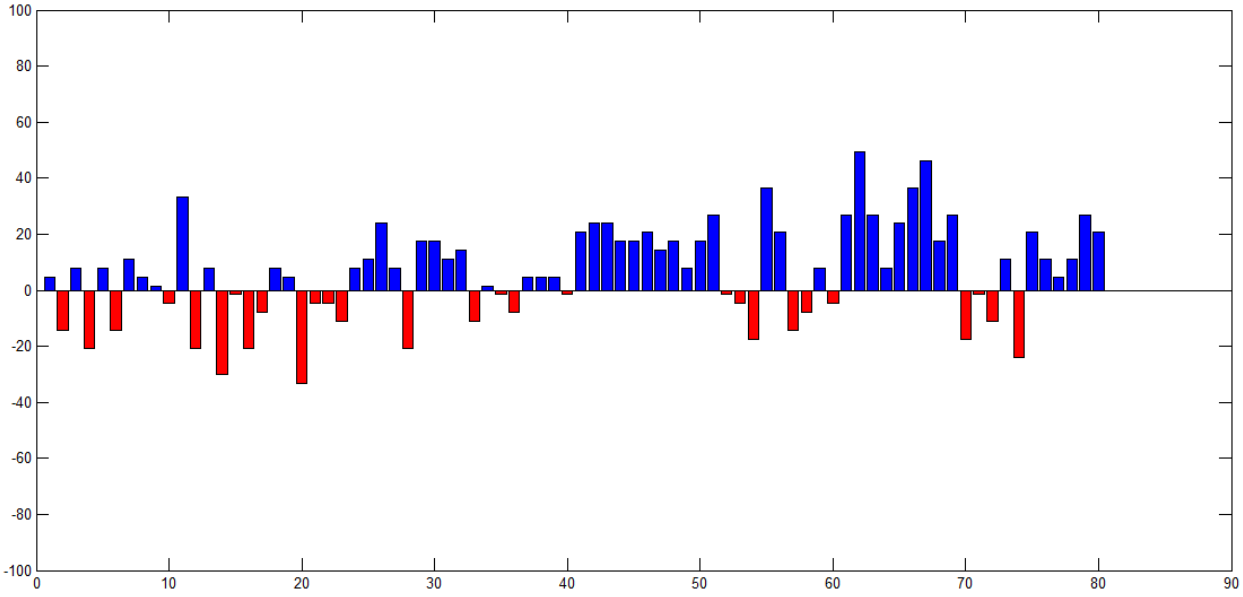

Figure 28: : classification of 80 epileptic and non-epileptic Channels with ICA feature reduction

	Sensitivity	Specificity	accuracy
Model 1 (Original)	97.5%	92.5%	95%
Model 2 (with PCA)	89.5%	42.5%	65%
Model 3 with (ICA)	80%	45%	62.5%

Table 4: Sensitivity and specificity of Models

3.3 Discussion

The main objective of this work is to efficiently perform an automated classification framework to detect seizure and non-seizure EEG signals. An EEG signals classification is proposed, which is based on DWT, then dimension reduction (PCA and ICA) and SVM classification. The procedures can be summarized as follows:

1- Wavelet decomposition was performed and the signals were divided into frequency sub bands. The frequency sub-bands of interest were in the range that contains the four signals $\alpha, \beta, \delta \; and \; \theta$ which is $0 - 30 \; Hz$.
2- Then statistical features were extracted from the decomposed signals and were used to create a feature matrix.
3- The number of features was high, for that dimensionality reduction using PCA and ICA was performed to obtain new data in a reduced space. This performed to remove the irrelevant features which are redundant and even degrade the performance of the classifier.
4- Classification process with support vector machine is carried out to identify epileptic and non-epileptic seizure. The SVM classifier used kernel function for more robust and simplification. A radial basis function (RBF) kernel was used. Also a cross validation of $kfold = 10$ were used to investigate the appropriate and the optimal kernel parameters $\sigma \; and \; \gamma$ to use in SVM training part.

In this study the training process was carried out using PCA + SVM and ICA + SVM. Subsei and Gursoy (2010) in their study made a comparison between PCA, ICA and LDA with SVM and Priya Balasubramanian in here thesis added a combination of ICA + PCA together to SVM.

The results shows that the SVM classifier performs better with original feature without reduction, by registering 97.5% sensitivity and 92.5% specificity and total accuracy of 95% . The time taken is considerably long compared to other models training. Priya.B gets the same results in her experiment and the training time was about 70 minutes which is a long time for processing data.

Performing PCA to SVM was able to correctly identify most of epileptic channels; sensitivity was about 89.5% and low specifity 42.5%. PCA was better than ICA in our study contrarily to the paper of (Subsei and Gursoy) who founds ICA accuracy 99.5% good compared to PCA 98.9%. Also they conclude with LDA as the best reduction and preferment technique, but has an execution time little bit long.

CONCLUSION

Conclusion

EEG signal analysis is gaining popularity in the field of neuroscience, brain-computer interface and physiological evaluation. EEG signal analysis most uses are in medical diagnosis especially for seizure detection.

An automatic classification technique of Support Vector Machine were implemented first with DWT decomposition and forwarded by a feature extraction than dimensionality reduction techniques PCA and ICA. As other studies results, The SVM classifier has superior classification capabilities compared to other classifiers, thus, it leads to the highest accuracy.

In this study an effective classification structure based on wavelet decomposition, feature extraction, PCA, ICA dimensionality reduction and SVM classifier were used on a set of data from five patients EEG records. Our results of EEG signal classification by using feature matrix has the highest accuracy than the reduced features. Nevertheless, the time of training process was too much long. PCA reduction gives less time execution and good sensitivity than ICA .The effectiveness of SVM detection needs a large scale of clinical trials EEG signal, as for that our results can differ from other papers and researchers studies. According to other papers, that use a large number of channels to train the SVM they founds that ICA leads to a good accuracy and it has a short time execution than PCA, and another paper states that LDA was the optimal among the others. To summarize, the application of this framework and SVMs can serve as a promising alternative for intelligent diagnosis system in the future.

Further approach of this work is to find another technique that gives a high performance classification, and also by choosing a best combination of analysis methods that leads to minimal time of training and gives high accuracy.

Bibliography

[1] Retrived from: iMotions. (2015, Sep.) IMOTIONS. [Online]. https://imotions.com/blog/top-3-devices-measuring-brain-activity/

[2] N. Jeremy Hill et al., "Recording Human Electrocorticographic (ECoG) Signals for Neuroscientific Research and Real-time Functional Cortical Mapping," *Journal of Visualized Experiment*, vol. 10, no. 64, pp. 3791-3993, Jun. 2012.

[3] J Satheesh Kumar and P Bhuvaneswari, "Analysis of Electroencephalography (EEG) Signals and Its Categorization - A Study," in *International Conference on Modeling, Optimization and Computing (ICMOC 2012): Procedia Engineering*, vol. 38, 2012, pp. 2525-2536.

[4] Rebeca Romo-Vàzquez, "Contribution à la détection et à l'analyse des signaux EEG épileptiques : débruitage et séparation de sources," Automatique, Traitement du signal et des Images, Génie informatique, Institut National Polytechnique de Lorraine, Doctorat Thése 2010.

[5] Retrived from: Macalester College: Private Liberal Arts College. [Online]. https://www.macalester.edu/academics/psychology/whathap/ubnrp/imaging/index.html

[6] P.Bhuvaneswari and J Satheesh Kumar, "Support Vector Machine Technique for EEG Signals," *International Journal of Computer Applications*, vol. 63, no. 13, pp. 1-5, Feb. 2013.

[7] Retrived from: (2017, Sep) The McGill Physiology Virtual Laboratory. [Online]. http://www.medicine.mcgill.ca/physio/vlab/biomed_signals/EEG_n.htm

[8] Erin Hecht and Dietrich Stout, "Techniques for Studying Brain Structure and Function," in *Human Paleoneurology*. New York: Springer Internationale Publisher, 2014, vol. 3, ch. 9, p. 250.

[9] Aashit K.Shah and Sandeep Mittal, "Invasive electroencephalography monitoring: Indications and presurgical planning," *Annals of Indian Academy of Neurology*, vol. 17, no. 5, pp. 89-94, Mar. 2014.

[10] Lung Chuin Cheong, Rubita Sudirman, and Siti Suraya Hussin, "Feature extraction of EEG signal using wavelet transform for autism classification," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 19, pp. 8533-8540, Oct. 2015.

[11] Quiroga Rodrigo Quian, "Quantitative analysis of EEG signals: Time-frequency methods and Chaos theory," Faculty of Engineering and Natural Sciences, Medical University, Lübeck,

PhD Diss. 1998.

[12] John L.Semmow, *Biosignal and Biomedical Image Processing - MATLAB Based Applications*. New York, U.S.A: Marcel Dekker, Inc., 2004, vol. 1.

[13] Retrived from: wavelet. [Online]. http://www.wavelet.org/tutorial/wbasic.htm

[14] Lei Lei, Chan Wang, and Xin Liu, "Discrete Wavelet Transform Decomposition Level Determination Exploiting Sparseness Measurement," *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, vol. 7, no. 9, pp. 1182-1185, Sept. 2013.

[15] Retrived from: Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension. [Online]. https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/

[16] Dominic Langlois, Sylvain Chartier, and Dominique Gosselin, "An Introduction to Independent Component Analysis:InfoMax and FastICA algorithms," *Tutorials in Quantitative Methods for Psychology*, vol. 6, no. 1, pp. 31-38, 2010.

[17] ALAN H.Fielding, *Cluster and Classification Techniques for the Biosciences*. New York: Cambridge University Press, 2007.

[18] Retrived from: Michael Halls-Moore. (2014, Sep.) QuantStart. [Online]. https://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners

[19] Abdulhamit Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659-8666, Dec. 2010.

[20] George Tzanetakis, George Essl, and Perry Cook, "Audio Analysis using the Discrete Wavelet Transform," in *Acoustics and Music Theory Applications*, 2001.

[21] Enamul Kabir, Siuly, and Yanchun Zhang, "Epileptic seizure detection from EEG signals using logistic model trees," *Brain Informatics*, vol. 3, no. 2, pp. 93-100, jun. 2016.

Appendix A: Matlab Code

```
if reset==1
    Var=loadEEG();
    Msignal(:,:)=Var(1:4096,:,2);
    Nsignal(:,:)=Var(1:4096,:,1);
   reset=0;
end
L=4096;
Fs=173.61;%23 seconds
t=(0:L-1)/Fs;
FM_inter=1:19;
PC_inter=20:24;
IC_inter=25:29;
%PC_IC=20:29;
Tp=63;
%*** normalization of datasets segment with mean=0 and stadar deviation=1
Msignal=zscore(Msignal,1,1);
Nsignal=zscore(Nsignal,1,1);
%****extracting feature matrix original and reduced with funtion
%compination for epileptic dataset
i1 = 1;
for i = 1:100;
i2 = i1+Tp-1;
[FM,PC,IC] = combination(Msignal(:,i),128,Tp);
%plot(perc)
%hold on
Mtrain(i1:i2,FM_inter) = FM;
Mtrain(i1:i2,PC_inter) = PC;
%Mtrain(i1:i2,IC_inter) = IC;
i1=i2+1;
end
%****extracting feature matrix original and reduced with funtion
%compination for epileptic dataset
i1 = 1;
for i = 1:100;
i2 = i1+Tp-1;
[FM,PC,IC] = combination(Nsignal(:,i),128,Tp);
Ntrain(i1:i2,FM_inter) = FM;
Ntrain(i1:i2,PC_inter) = PC;
Ntrain(i1:i2,IC_inter) = IC;
i1=i2+1;
end
 %******** using 40 first channels for testing the SVM and the last
 %channels for training

%Mtest1=Mtrain(1:45*Tp,FM_inter);
Mtest=Mtrain(1:40*Tp,PC_IC);
Mtrain=Mtrain(40*Tp+1:end,PC_IC);

%Ntest1=Ntrain(1:45*Tp,FM_inter);
Ntest=Ntrain(1:40*Tp,PC_IC);
Ntrain=Ntrain(40*Tp+1:end,PC_IC);

test=[Mtest;Ntest];
```

```
%SVM training
if start==1
    %*** buld the training matrix from 60 channels epileptic and 60
    %channels non-epileptic
data=[Mtrain(1:60*Tp,1:10);Ntrain(1:60*Tp,1:10)];%48 channel M x 32 channel N
theclass=ones(size(Mtrain,1)+size(Ntrain,1),1);%2832 row =-1, 1888 row=1;
theclass(1:size(Mtrain,1))=-1;
%**** Mtrainc and Ntrainc are training matrices for cross-validation
Mtrainc=Mtrain(1:5*Tp,1:10);
Ntrainc=Ntrain(1:5*Tp,1:10);
datac=[Mtrainc;Ntrainc];
classc=ones(10*Tp,1);
classc(1:5*Tp)=-1;
 %**** selecting the optimal value of z and used as parameter for training
 %the svm
c=cvpartition(size(datac,1),'Kfold',10);
minfn= @(z) crossval('mcr',datac,classc,'Predfun',...
    @(xtrain,ytrain,xtest)crossfun(xtrain,ytrain,xtest,...
    exp(z(1)),exp(z(2))),'partition',c);

opts=optimset('TolX',5e-4,'TolFun',5e-4);

%**** the oprimal value of z o reach must be near the unit value [1;1]
z=zeros(2,1);
i=1;
tic
while (z(1)<0.6 || z(1)>1.6)||(z(2)<0.6 || z(2)>1.6)

[searchmin,fval] = fminsearch(minfn,randn(2,1),opts);
z=exp(searchmin)
end

%*** SVM training
svmStruct=svmtrain(data,theclass,'Kernel_Function','rbf',...
    'rbf_sigma',z(1),'boxconstraint',z(2));
start=0;
end

%***** SVM classifing
i1=1;
for i=1:80
    i2=i*Tp;
group=svmclassify(svmStruct,test(i1:i2,:));
i1=i2+1;
test1(:,i)=group;
end
toc
```

```
x=zeros(80,1);

for i=1:80
    dif=sum(test1(:,i));
    x(i)=dif*100/Tp;
end
%{...%}

%plot the (+1) with blue and (-1) with red
y=zeros(80,1);
h=zeros(80,1);
for j=1:80
    if x(j)>0
    y(j)= x(j);
    else
        h(j)=x(j);
    end
end
figure(1)
bar(y,'b');ylim([-100,100]);
hold on
bar(h,'r');ylim([-100,100]);
%}
```

Appendix B: Matlab Functions

```
function yfit =crossfun(xtrain,ytrain,xtest,rbf_sigma,boxconstraint)
svmStruct=svmtrain(xtrain,ytrain,'Kernel_Function','rbf',...
    'rbf_sigma',rbf_sigma,'boxconstraint',boxconstraint);
yfit=svmclassify(svmStruct,xtest);
end
```

```
function [F_Matrix,PCA_matrix,ICAFinal,explained,perc]=combination(signal,WL,Tp)
[A5,D1,D2,D3,D4,D5]=Wavelet_Decomposition(signal,5);
%*******************************
 %feature extraction
 %*******************************

 OL=WL/2;
 %Statistical features for wavelet subband D2
[MeanD2,AvgD2,SDD2] = Feature_ex(D2,OL,WL);
%Statistical features for wavelet subband D3
[MeanD3,AvgD3,SDD3] = Feature_ex(D3,OL,WL);
%Statistical features for wavelet subband D4
[MeanD4,AvgD4,SDD4] = Feature_ex(D4,OL,WL);
%Statistical features for wavelet subband D5
[MeanD5,AvgD5,SDD5] = Feature_ex(D5,OL,WL);
%Statistical features for wavelet subband A5
[MeanA5,AvgA5,SDA5] = Feature_ex(A5,OL,WL);
% Calculating the ratios R1,R2,R3 and R4
RatioR1 = rdivide(MeanD2,MeanD3);
RatioR2 = rdivide(MeanD3,MeanD4);
RatioR3 = rdivide(MeanD4,MeanD5);
RatioR4 = rdivide(MeanD5,MeanA5);
 %time_point_number
Fn=19;
F_Matrix=zeros(Tp,Fn);
F_Matrix(:,1)=MeanD2';
F_Matrix(:,2)=MeanD3';
F_Matrix(:,3)=MeanD4';
F_Matrix(:,4)=MeanD5';
F_Matrix(:,5)=MeanA5';
F_Matrix(:,6)=AvgD2';
F_Matrix(:,7)=AvgD3';
F_Matrix(:,8)=AvgD4';
F_Matrix(:,9)=AvgD5';
F_Matrix(:,10)=AvgA5';
F_Matrix(:,11)=SDD2';
F_Matrix(:,12)=SDD3';
F_Matrix(:,13)=SDD4';
F_Matrix(:,14)=SDD5';
F_Matrix(:,15)=SDA5';
```

```
F_Matrix(:,15)=SDA5';
F_Matrix(:,16)=RatioR1;
F_Matrix(:,17)=RatioR2;
F_Matrix(:,18)=RatioR3;
F_Matrix(:,19)=RatioR4;

perc=zeros(19,1);
%Correlation = corr(F_Matrix,F_Matrix);
[wcoeff,score,latent,tsquared,explained] = pca(F_Matrix,...
'VariableWeights','variance');
coefforth = (diag(std(F_Matrix)))\wcoeff;
%figure(1);
%plot(explained);
%title('non-healthy patient');
%xlabel('Principal Component');
%ylabel('Variance (%)');
for j=1:19
   perc(j,1)=sum(explained(j:19))*100/sum(explained);
end
%plot(perc);
PCA_matrix(:,:)= score(:,1:5);

NumberIC = 5; % number of Independant components we are interested in
FeatureICA = F_Matrix';
% Cumputing the ICA using the jadeR function
W = jadeR(FeatureICA,NumberIC);
ICAFinal = (W*FeatureICA)';
end
```

```
function [ A5,D1,D2,D3,D4,D5] = Wavelet_Decomposition(S,L)
%UNTITLED3 Wavelet decomposition of the original signal
% C is the vector formed by concatenating approximation and detail
% coeeficients at each level. L is the vector that gives the length of each
% component. Courtsey:
% http://www.mathworks.com/help/wavelet/ug/one-dimensional-
% discrete-wavelet-analysis.html#f4-997029
% S = signal, L = level of the decomposition.

[C,L] = wavedec(S,L,'db6');
cA5 = appcoef(C,L,'db6',5); % Extracts approximation coefficents
% Extracts detailed coefficients
[cD1,cD2,cD3,cD4,cD5] = detcoef(C,L,[1,2,3,4,5]);
% Reconstructs the signal component corresponsing to each of the six
% wavelet coefficient sequences
A5 = wrcoef('a',C,L,'db6',5);
D1 = wrcoef('d',C,L,'db6',1);
D2 = wrcoef('d',C,L,'db6',2);
D3 = wrcoef('d',C,L,'db6',3);
D4 = wrcoef('d',C,L,'db6',4);
D5 = wrcoef('d',C,L,'db6',5);
end
```

```
function [ Mean,Avg,SD ] = Feature_ex( SB,OL,WL )
%Extracts the statistical features of a subband
% 1. Mean of the absolute values of the coefficients in each sub band
% eg. If there are 20000 samples so 20s worth of data, the windown length
% for the statistical analysis will be 0.5s each with an overlap of 0.25
% seconds and that will give a total of 79 time points or sample points
% for each subband.
% 2. Average power of the wavelet coefficients in each sub band
% 3. Standard deviation of the coefficients in each sub band
SignalBlock = (length(SB)/OL)-1;
SignalEnd = WL - 1;
% Preallocating matrix for mean, avg and std
Mean = zeros(1,SignalBlock);
Avg = zeros(1,SignalBlock);
SD = zeros(1,SignalBlock);
i1 = 1;
for i = 1:SignalBlock
i2 = i1+SignalEnd;
%Statistical features for the wavelet subband
Mean(i) = mean(abs(SB(i1:i2)));
Avg(i) = mean((SB(i1:i2).^2));
SD(i) = std((SB(i1:i2).^2));
i1 = (i2-OL)+1;
end
end
```

```
function [Var]=loadEEG()
labindex=[1:100];
for j=1:5
p=['dataset\O';'dataset\S';'dataset\N';'dataset\F';'dataset\Z'];
ext='.txt';
name2=['O';'S';'N';'F';'Z'];
name1=['O0';'S0';'N0';'F0';'Z0'];
name=['O00';'S00';'N00';'F00';'Z00'];
 p=p(j,:);
 name=name(j,:);
 name1=name1(j,:);
  name2=name2(j,:);
for i=1:100
if i>9
    if i<100
        name=name1;
    else
        name=name2;
    end
end
filename = fullfile(p,[name num2str(labindex(i)) ext]);
Var(:,i,j)=load(filename);
end

end
end
```