Analysis of EEG Signals for Detection of Epileptic Seizure

Author

Saqib Ejaz Awan

NUST201362469MCEME35213F

Supervisor

Dr. Muhammad Usman Akram

Co Supervisor

Dr. Muazzam Ali Khan

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

FEBRUARY, 2015

Author

Saqib Ejaz Awan

NUST201362469MCEME35213F

A thesis submitted in partial fulfilment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor:

Dr. Muhammad Usman Akram

Thesis Supervisor's Signature:_____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

FEBRUARY, 2015

Declaration

I hereby certify that I have developed this thesis titled *"Analysis of EEG signals for detection of Epileptic Seizure"* entirely on the basis of my personal efforts under the sincere guidance of my supervisor Dr. Muhammad Usman Akram. All of the sources used in this thesis have been cited properly and contents of this thesis have not been plagiarized. No portion of the work presented in this thesis has been submitted in support of any application for any other degree of qualification to this or any other university or institute of learning.

Signature of Student

Saqib Ejaz Awan

NUST201362469MCEME35213F

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Saqib Ejaz Awan

NUST201362469MCEME35213F

Signature of Supervisor

Dr. Muhammad Usman Akram

Copyright Statement

- Copyright in text of this thesis rests with the student author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the author and lodged in the Library of NUST College of E&ME. Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.

- The ownership of any intellectual property rights which may be described in this thesis is vested in NUST College of E&ME, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the College of E&ME, which will prescribe the terms and conditions of any such agreement.

- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST College of E&ME, Rawalpindi.

Acknowledgments

I am indebted to a lot of people on the successful completion of this research but first and foremost I am thankful to ALLAH for his countless blessings and for giving me the strength and courage to undertake this novel project and see to its end. I am eternally grateful to my family whose support and prayers helped me achieve my aim and more. I'd also like to thank my supervisor Dr Usman Akram who nudged me in this direction and laid rest to my qualms. Lastly, this project would never have seen the light of day had it not been the constant guidance and backing of my supervisor Asst. Prof Sajid Gul Khawaja, who was instrumental throughout the development of this work and for his constant help whenever I asked for it. Thank you sir for everything, it was truly an honour learning under your tutelage.

Abstract

Epileptic seizures are known to be sudden surges of electrical activity in the brain which cause the affected person to behave abnormally for a short time period. The human brain produces electrical signals which prove vital in understanding the degree of abnormality that may, in many cases, result in a person behaving unusually. The information contained in these signals is recorded via an EEG machine, which is able to extract even the most subtle details from the electrical waves that the brain signals generate. Usually, the signals from the aforementioned device are interpreted by the specialists who specialize in this very thing but their detection is susceptible to errors which prove fatal in some cases. This research presents an autonomous system, capable of detecting the occurrence of an epileptic seizure, without the help of an expert. The proposed system consists of four steps i.e. pre-processing, feature extraction, feature selection and classification. The purpose of pre-processing is to organize the data in an orderly manner and to remove noise. We have also applied Laplacian smoothing on multichannel data to generate a surrogate channel having information of all channels. The feature extraction phase extracts temporal, spectral and time-spectral domain features for proper representation of seizure and non-seizure samples. The system then performs the process of feature selection, where the best set of features are determined using rank features and are finally used for classification of EEG signals as normal or abnormal using a hybrid classifier. The proposed system is tested on a publicly available dataset and results show the significance of the proposed system.

Table of Contents:

List of Figures:

List of Tables:

List of Abbreviations:

ApEn	Approximate Entropy
ANN	Artificial Neural Networks
ASCII	American Standard Code for Information Interchange
DWT	Discrete Wavelet transform
EDF	European Data Format
EEG	Electroencephalography
FFNN	Feed forward Neural Network
GMM	Gaussian Mixture Model
HFD	Higuchi Fractal dimension
IApE	Improved Approximate Entropy Method
IMFs	Intrinsic mode functions
LBDWT	Lifting-based discrete wavelet transform
LDA	linear discriminant analysis
LR	logistic regression
MEG	Magneto Encephalography
MLPNN	multilayer perceptron neural network
MWT	Multi-Wavelet Transform
NCOV	normalized coefficient of variation
PFD	Petrosian Fractal Dimension
PNN	Probabilistic Neural Network
RIR	Relative Intensity Ratio
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine#
TFDs	t-f distributions

Chapter 1:   Introduction

Brain controls the complete human body. When humans do any activity either physical or just thinking or feeling, brain generates signals to accomplish these tasks. These signals are electrical and chemical signals communicated between the neurons present in the brain and the rest of the human body. The electrical signals can be measured with the help of a technique called Electroencephalography (EEG). In EEG, a number of electrodes are attached to the human scalp. This is mainly done to catch the disorders of human brain as seen by the difference in electrical signals when compared to a normal brain signals. EEG may reveal dementia, coma, seizure, narcolepsy and other brain problems. It is a harmless process with the purpose of just measuring the electrical pulses. There are other techniques available for this purpose as well. One such technique is Magnetoencephalography (MEG) which maps brain activity by recording magnetic fields produced by electrical currents which occur naturally in the brain, using very sensitive magnetometers. This technique, however, is very expensive in comparison to EEG and difficult to comprehend as well.

Seizure is a change in behaviour of a person after an abnormal electrical activity occurs in the human brain. It may appear as convulsion in which a person's body shakes rapidly without the control of the individual. For some people it occurs once or twice, while for others, it may rise to thousands per day. Having a seizure does not necessarily mean that a person has epilepsy. Two or more seizures at one point in time is taken as epilepsy. Epileptic seizures are neither contagious nor a result of mental illness.

Any irregular neuron activity of the brain may cause a seizure. However, a seizure rarely damages the brain, but it may. Around half the number of seizures have no definite cause. Epilepsy, on the other hand, has several reasons for its occurrence. A research reveals that the membrane which is attached to the neurons of human brain has a major role in epilepsy. Genetic abnormalities also contribute towards the epilepsy. According to [1], epilepsy is the most common disorder of brain and affects around 1% of the total population of the world. Epilepsy is broadly divided into partial and generalized seizures. Partial seizure is seen in only a few channels attached to the scalp but generalized seizure is reflected in all the channels providing a variation in measure of electrical activity with the help of attached electrodes. Figure 1.1 reflects a seizure in an EEG recording, seen as spikes (impulsive fluctuations) [2].
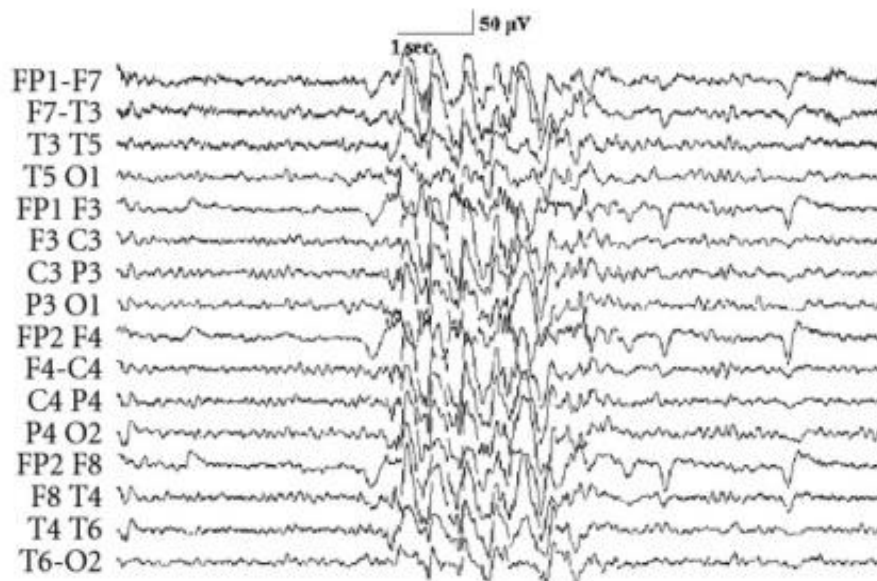
Figure 1.1 Occurrence of seizure in EEG

There are a variety of algorithms proposed to detect an epileptic seizure. This report reveals an automated system equipped with the capability of applying certain signal processing methods for detection of an epileptic seizure. It has four major parts: pre-processing, feature extraction, feature selection followed by classification in the end. Figure 1.2 shows a flow diagram of the proposed automated system.

1.1 Motivation

Epileptic seizures are considered as a major brain disorder which affects a big 1% of the entire world population and around 0.2% of the affected individuals lose their lives when epilepsy is poorly controlled. Patients sometimes lose memory function, in 10 years follow-up, when not provided with proper epileptic surgeries. The injuries normally occur due to a lack of awareness of the individuals regarding epileptic seizures or a poor communication exchange between patient and the specialist. Pakistan, according to statistics, lacks both in the trained specialists as well as resources when it comes to automated diagnostic system for epileptic seizure detection. About 9.99 per 1000 Pakistanis suffer epilepsy. A majority of this ratio is younger than 30 years. A large number of people from rural areas suffer this disease as compared to urban areas. A study reveals that 1.38 million people suffer from epilepsy in our country with the number of trained neurologists is less than 30. This leads to a very large number of patients per specialist which is the reason why so many people are not provided proper treatment in

case of epileptic seizures. This leads to a very large number of patients per specialist which is the reason why so many people are not provided proper treatment in case of epileptic seizures. The situation demands the awareness of the epilepsy to be encouraged. With all this in mind, we aim to develop a system, automated in nature, for the diagnosis of epileptic seizures.
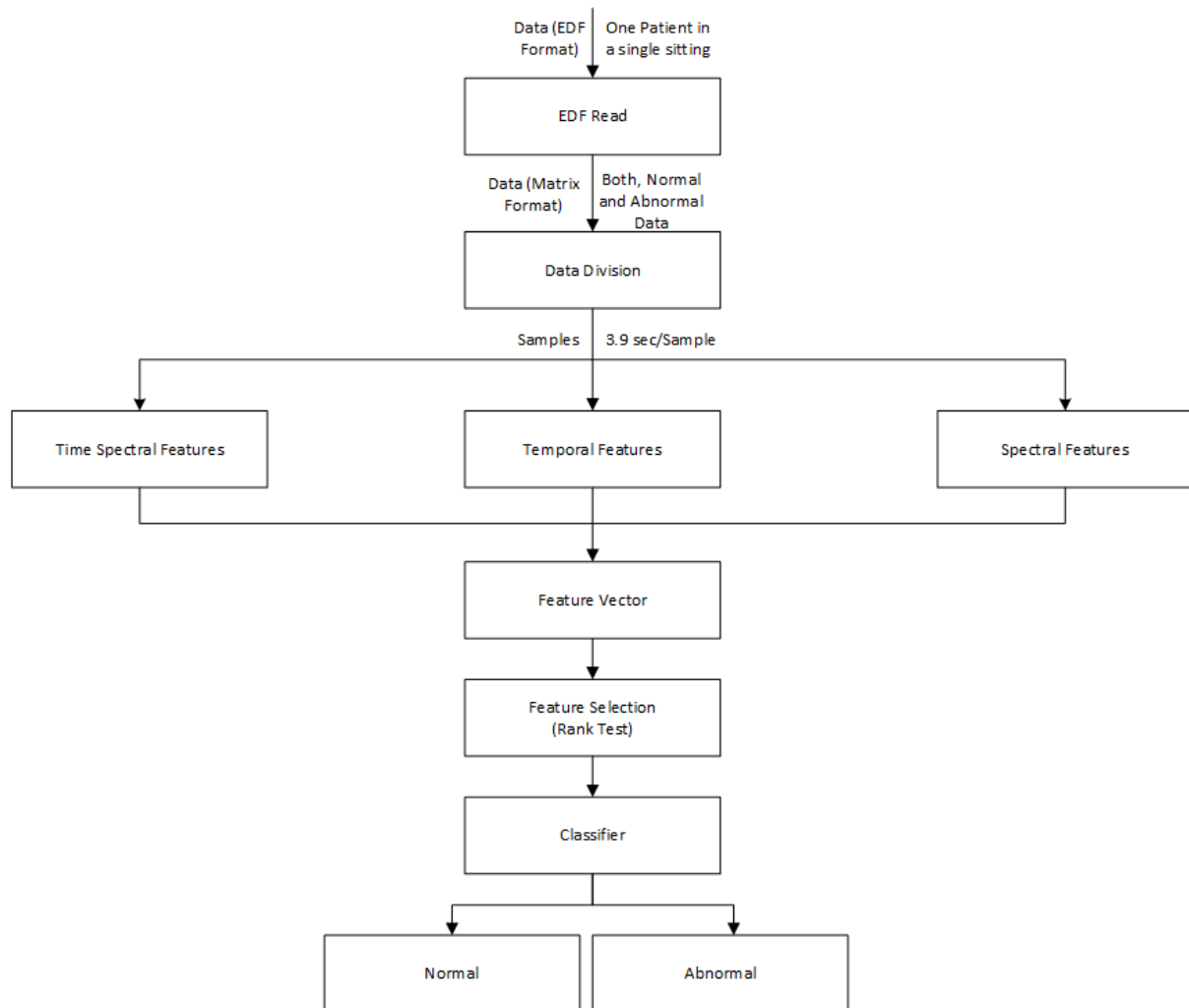
Figure 1.2 Flow Diagram of proposed system

Figure 1.2 shows the basic level breakdown of this research. The EEG recorded from the scalp is given as input to the system which break downs the data into chunks. Then features of different domains are computed out of the data. After that a feature vector is formed. Then the number of features present in the feature vector is reduced to the minimum keeping the best accuracy under consideration. This step is followed by the classification step which determines a signal as either Normal or Abnormal.

1.2 Scope & Objectives:

The epilepsy diagnostic procedure is aided by the study of EEG, which is the prime issue we are lacking in our country. So research in the EEG domain will not only benefit the entire diagnostic system but it will also bridge the gap between medical industry and the research field. Besides developing precise methods for automated detection of seizures, this research has the following objectives.

- To propose and produce an intelligent and automated detection system for classification of seizure.
- To save the time of neurologists during evaluation process of EEG waves for detection of seizures.
- To create a research environment in the field of biomedicine in the educational institutes of Pakistan.
- To create awareness in people regarding epilepsy and its effects, especially in students.
- To emphasise the need of a proper diagnostic system in Pakistan.
- To design algorithms for cost effective diagnosis for people suffering from epilepsy.

1.3 Challenges

The traditional method of determining epilepsy is the continuous monitoring of EEG recordings which is both tedious and time consuming. We want to automate this procedure by developing a computer aided automated diagnostic system A variety of methods are already present for the diagnosis of epileptic seizures, which still have room for improvement. Some of the challenges that are relevant to our proposed system are given below:

1.3.1 Pre-processing of data

The data used in this research is from the CHB-MIT scalp EEG database. The first challenge is to pre-process the database to make it suitable to handle and remove the noise present in it. The de-noised data has been divided into blocks of 3.9 seconds of EEG recordings. This is done by analysing the occurrence of an event in a certain EEG sample.

1.3.2 Feature Extraction

Feature extraction is the second challenge in this research. A number of features that can be used for the detection of epilepsy are present. The feature extraction phase is meant to extract temporal, spectral and time-spectral domain features for proper representation of seizure and non-seizure samples.

1.3.3 Feature Selection

One of the biggest challenges of our research is to determine a set of best features from a pool of available features to improve our results. We are using, Wilcoxon signed - rank test [4] and the Ansari-Bradley test [5] to achieve the mentioned goal. We are also applying Laplacian smoothing on multichannel data to generate a surrogate channel having information of all channels.

1.3.4 Classification

Another challenge in this research is to classify the test data as normal or abnormal. Support Vector Machine (SVM) is used for the purpose of classification. A set of labelled training data is fed to the classifier and the system is trained to detect any test sample as either normal or abnormal.

1.4 Structure of Report

Rest of the report bears the following format:

Chapter 2 explains epilepsy in detail, its causes and types. It gives details about how epilepsy affects the patients, what is electroencephalography (EEG) and how it is acquired.

Chapter 3 presents an overview of the previously designed algorithms and methods for the detection of epileptic seizures and summarizes certain algorithms and their accuracies.

Chapter 4 is the main body of this report since it explains the how pre-processing, feature extraction, feature selection and classification are actually done in this research.

Chapter 5 contains all the tests and their results.

Chapter 6 covers the conclusion and the future work for the proposed automated system.

1.5 Summary

A large number of people around the globe suffer epileptic seizure due to abnormal electrical activity of the brain neurons. This electrical activity can be captured with the help of electroencephalography. So, an automated system, capable of detecting epileptic seizure using EEG signal, various signal processing and machine learning techniques, can be designed. The overall system is composed of blocks of pre-processing, feature extraction, feature selection and classification.

Chapter 2: Epileptic seizure

This chapter gives a brief overview of an epileptic seizure from the medical perspective. It explains causes, signs and different effects related to epilepsy. Furthermore it contains different treatment methods or therapies based on different algorithms using EEG.
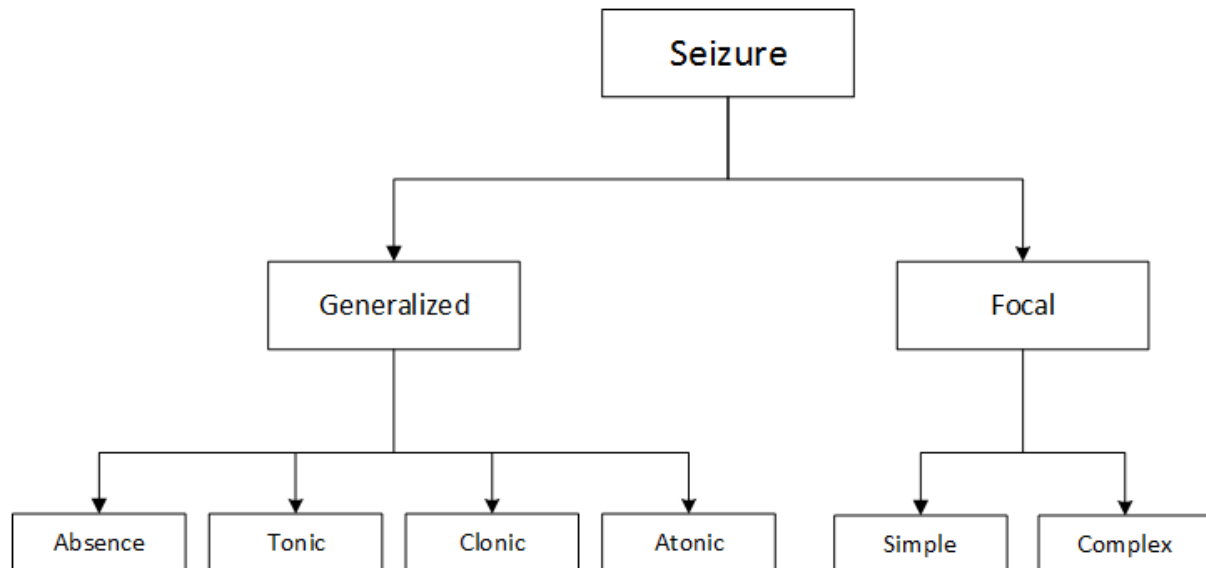
2.1 Epilepsy

2.1.1 Introduction

Epilepsy is a general term for multiple conditions where two or more unprovoked seizures are observed. A seizure occurs due to sudden electrical disturbance in the brain. Seizure is considered the second most powerful neurological disorder after stroke. Until the 19th century, epilepsy was not acknowledged as a brain disorder. The main cause of epileptic seizure is genetic disorders, strokes and brain tumours though the causes of seizures are sometimes non-deterministic in nature. A seizure occurs when the brain cells send out wrong signals. The difference between seizure and epilepsy must be clear to all. We can define epilepsy as a series of repetitive seizures. Seizures may cause damage to brain but most of them don't have an effect on the brain. Seizures may occur at any age with duration ranging from a few seconds to a few minutes.

Seizures can be broken down into different categories. This division is based on the brain area which is affected as a result of seizure. Seizures are broadly classified as focal seizures and generalized seizures. Focal seizures or partial seizures occur in a certain part of the brain not the complete brain. In a simple focal seizure, a person remains conscious but experiences an unusual feeling of joy, anger or sadness. However in a complex focal seizure, the person loses his/her consciousness. Figure 2.1 shows a simple seizure classification pattern.

Seizures occurring in both sides of the brain is termed as generalized seizure. Absence seizure, tonic seizure, clonic seizures and atonic seizures are few types of generalized seizures. There is a fairly thin boundary between focal and generalized seizures. Sometimes initial focal seizures spread over the entire brain to form generalized seizure. Some people may have both types of seizures but with no clear pattern between them. Statistics guide that 60% of the epileptic patients have seizures in which body muscles contract and expand rapidly without the intervention of human brain. Such seizures are called convulsive seizures. There is equal ratio of epileptic patients having less than one seizure per year to number of seizures between one and twelve seizures per year and more than one seizure per month [6]. Figure 2.1 below shows

the brief categorization of seizures as Generalized and Focal seizure. They are further broken down into many types as mentioned below.

Figure 2.1 Seizure Classification

2.1.2 Causes of Epilepsy

The number of epilepsy cases reported each year is about 1, 80,000 out of which 30% are children. Different causes of epilepsy are observed in different age. The main cause of epileptic seizure is genetic disorders, strokes and brain tumours though the causes of seizures are sometimes non-deterministic in nature. Around 30% people catch seizure owing to change in the structure of brain. Autism spectrum disorder in children leads to epilepsy however its causes are still to be determined. Lack of oxygen in new-borns and infections in case of children and infants also becomes a reason for epilepsy.

Epilepsy is categorized as following: symptomatic epilepsy, idiopathic epilepsy and cryptogenic epilepsy. Some specific causes like drugs, alcohol and birth abnormalities may also cause Symptomatic epilepsy. Idiopathic epilepsy occurs due to no specific cause whereas a brain damage may result in a cryptogenic epilepsy. Epilepsy can also occur as a result of abnormal level of sodium, sugar or blood.

Seizures and epilepsy can be avoided by evading the following:

- Use of alcohol
- Drug abuse

- Lack of proper sleep
- Anxiety / Stress

Epilepsy adversely affects the patient's health. Anxiety, depression, loss of memory and mental injury are all the signs of this disease.

2.2 Electroencephalography (EEG)

2.2.1 Introduction

Human brain has two parts: left hemisphere and right hemisphere. The left hemisphere has the control of right side of body while the right hemisphere controls the left side of the body. A portion of brain is present to control vision and some part of brain controls the memory. The neurons present in the brain produce certain electrical and chemical signals to perform certain functions. These signals are termed as electroencephalogram (EEG). When we record these signals it is called EEG.

There are four lobes in each hemisphere, each one described below:

Frontal lobe is located at the front of both the hemispheres. The frontal lobe predicts the future from the results of the current state. It also helps maintaining the long term memory. Parietal lobe is located just behind the frontal lobe. This lobe integrates sensory information from all the parts of body. Temporal lobe is part of the cerebral cortex. It is located on the lower side of each hemisphere. It has been assigned functions like long term memory and emotional responses. The occipital lobe is the smallest amongst all. It is located in the rearmost portion of the brain. It contains areas of vision. Figure 2.2 shows the brain's structure and all the four lobes [14].

The electrodes used in the process of EEG govern which type of recording is to be examined. There are two types of EEG readings: scalp EEG and intracranial EEG. Scalp EEG generates a signal which may carry noise due to the effect of tissues. EEG is considered as the primary tool to diagnose epileptic seizures. Usually different frequency bands are observed where each band is linked to a special activity of the human body. These bands and their activities are elaborated below.

The slowest activity, though highest in amplitude, is called Delta. Its frequency is 3 Hz or below. It is prominent in the frontal lobe in adults and babies. The slow activity is called Theta. Its frequency ranges from 3.5 to 7 Hz appears in normal children, aged up to 13 years, in sleep. The Alpha activity appears upon closing eyes / relaxing and disappears upon opening eyes /

alerting. Its frequency is normally 7.5 to 13 Hz. The fast activity with frequency range of 14 to 30 Hz is called Beta activity. It is usually found absent in the area of damaged brain. Gamma activity governs motor functions. Its frequency ranges from 30 to 100 Hz.
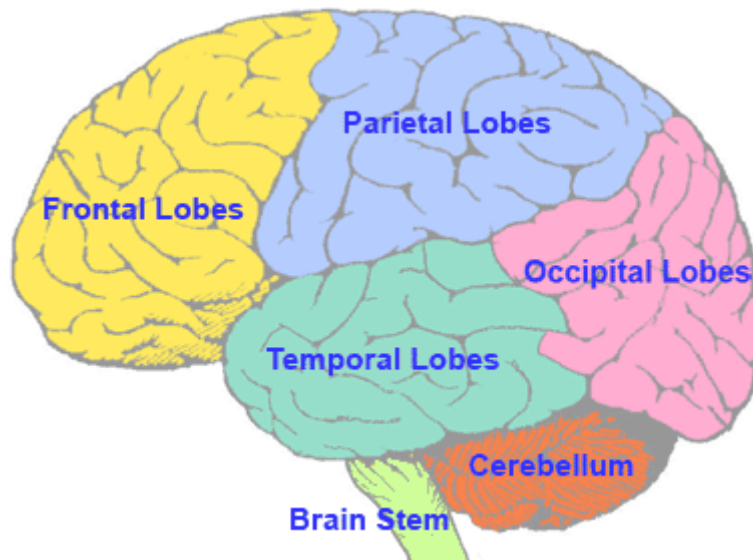
Figure 2.2 Structure of Brain and its lobes

2.2.2 EEG Acquisition
The prime constituents of an EEG acquisition are as follows:

1. Electrodes

2. Amplifiers

3. Filters

4. Analogue to digital converter

5. A recording device

Electrodes are meant to read electrical signals from scalp. As the signals captured by the electrodes are fairly week, we use amplifiers which amplify those microvolt signals to suitable strength for the next step of digitization. The purpose of analogue to digital converter is to convert the signal into digital format which allows it to display on the screen. The scalp recordings are a measure of the charge which flows through the active electrodes with respect to a reference electrode. Figure 2.3 shows an epilepsy patient bearing electrodes over his scalp [15].
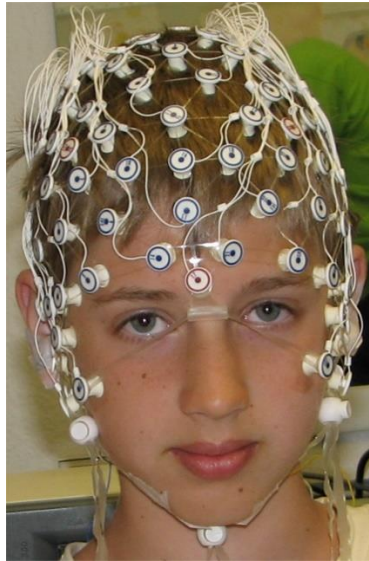
Figure 2.3 Electrode Cap

Electrodes: Electrodes are the most critical components of the mechanism of obtaining EEG. Their placement over the scalp is also of key importance. There are plenty of types of electrodes each having its own features and characteristics. Few are mentioned below:

1. Re-useable disc electrodes
2. Disposable electrodes
3. Cap Electrodes
4. Needle electrodes

A conducting paste is applied on the scalp on precise electrode positions. It results in a stable connection and reduced resistance between the scalp and electrodes.

Electrode Placement: The 10-20 electrode positioning system is one of the many electrode placement methods we have in market. Electrode placements are labelled below in the figure where different areas of brain are represented as frontal (F), central(C), temporal (T), posterior (P) and occipital (O) [13].

According to the 10-20 placement system, the distance between two adjacent electrodes is kept either 10% or 20% of the total front – back or right – left distance of the skull which is why this system is named as the 10 – 20 placement system. Figure 2.4 shows the placement of electrodes over the scalp according to 10-20 electrode system. Each lobe is associated with a letter while the number represents the location of hemisphere. There exists no central lobe. Cz is mentioned merely for identification purposes. Even numbered electrodes represent the

electrodes place on the right hemisphere whereas the odd numbered electrodes represent the left hemisphere's electrodes.
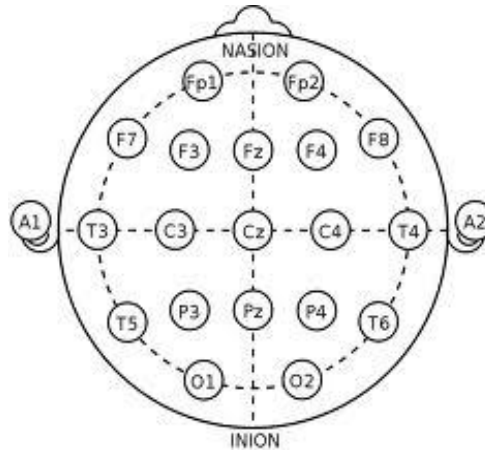
Figure 2.4 10-20 electrode placement system

2.3   EEG - MEG Comparison

MEG is another technique used for the diagnosis of epileptic seizures. It is a test which utilizes sensors to generate an image representation of the magnetic fields produced by the human brain. Doctors take help of these images to determine if a patient carries epilepsy or not. A comparison of EEG and MEG is given in the table 2.1.

S/No	Characteristics	EEG	MEG
1	Attenuation	more attenuation by scalp	less attenuation by scalp
2	Signal strength	10mV	10fV
3	Dipole position	Tangential or radial	Tangential only
4	Cost	Less expensive	More expensive than EEG
5	Overall performance	Low	High

Table 2.1 Comparison of EEG and MEG

The table shows that EEG is less expensive than MEG however it captures more attenuation by scalp as compared to MEG. The overall performance of EEG is lower than that of MEG. However, the dipole position captured by EEG can be tangential or radial contrary to MEG which has only tangential dipole position and if proper signal processing techniques are applied over the EEG signals, it can produce a working system with relatively lower cost.

2.4   Summary

This chapter presented a detailed overview of epilepsy and how to detect epilepsy. Epilepsy is termed as a group of conditions, primarily, more than one unprovoked seizures in a limited amount of time due to irregular electrical activity of brain neurons. This activity is captured with the help of an EEG machine. The mechanism of EEG acquisition is governed by different electrode placement techniques which in this study is the 10-20 electrode placement method.

Chapter 3: State of the art techniques for Epileptic seizure diagnosis

EEG plays an important role in the diagnosis of diseases like epilepsy, tumours and stroke. The main process of epilepsy detection is the classification of EEG data as normal or abnormal. Different algorithms can be used for this purpose. This chapter contains a brief review of the performance of some specific algorithms.

3.1 Literature Review

Several algorithms have been proposed for the automation of epileptic seizure detection. Given below are the reviews of some of the already proposed methods and their efficiency.

3.1.1 Multi-Stage Fuzzy Rule-based Algorithm

In [16], the authors proposed a multi – stage fuzzy rule-based algorithm to enable the detection of epileptic seizures. Features used to implement the proposed technique consisted of average amplitude, rhythmicity (coefficient of variation of amplitude), entropy and dominant frequency. In order to combine all features, an adaptive fuzzy sub-system was designed, the output of which was used in a threshold procedure to determine the final result. iEEG datasets taken out from Freiburg Seizure Prediction EEG (FSPEEG) database were used to test the designed system, which yielded 95.8% accuracy, with an average detection latency of 15.8 seconds and a false detection rate of 0.26 per hour.

3.1.2 Multi-wavelet Transform, Neural Network and Enhanced Approximate Entropy

Sharanreddy et al. [17], presented a novel technique, combining multi-wavelet transform, feed forward artificial neural network and enhanced approximate entropy, to facilitate signal classification for epilepsy seizure detection. EEG signal was first converted from the EDF (European Data Format) to ASCII (American Standard Code for Information Interchange) format, which was then provided as input to MWT (Multi-Wavelet Transform). Irregularities were then determined in the brain signal, using IApE (Improved Approximate Entropy Method). IApE was then trained using a feed forward neural network (FFNN). CHB-MIT Scalp EEG Database was used as the testing database. Approximate accuracy of 90% was achieved. Another innovative method for the automatic detection of seizure onset was presented in [18]. It made use of two statistical features: skewness and kurtosis, and a wavelet based feature: normalized coefficient of variation (NCOV), and a simple linear classifier. The proposed system was evaluated, using data for 10 patients, from the CHB-MIT scalp EEG database. A

mean latency of 3.2 seconds was observed, with a mean false detection rate of 1.1 (per hour) and 100% sensitivity.

In [19], Ali Shoeb et al. evaluated a machine learning approach to construct patient specific classifiers. Features used include rhythmic activity, channel identity and a non-EEG feature (such as ECG) to ascertain the final output. The methodology was tested on the CHB-MIT database and yielded a median false detection rate of 2 false detections per 24 hours.

3.1.3   Probabilistic Neural Network

[20] discussed a new approach for epileptic seizures detection, based on Probabilistic Neural Network (PNN). After the process of data acquisition, Discrete Wavelet Transform (DWT) was used to derive wavelet coefficients. Classification was based on the frequency domain feature of EEG, known as "Approximate Entropy" - (ApEn). Datasets used were taken from the NIMHANS, Bangalore database. For generalized epilepsy, 90% accuracy was achieved. For ECT epilepsy, 73%, and for detection of healthy signals, 68% accuracy was achieved.

The authors in [21] explored a method for automatic epilepsy identification, with the aid of 'interictal' scalp EEG, along with a modified approach to PNN. Features used included Relative Intensity Ratio (RIR), Petrosian Fractal Dimension (PFD), Higuchi Fractal dimension (HFD) and Hjorth Parameters. These features were then fed into the PNN. To better classification correctness, a component classifier for each channel was built, and the final classification decision was determined by the combined vote of each component classifier. The dataset used was obtained from the Dept. of Neurosurgery, Jiangsu Provincial Hospital of Chinese Medicine, China. This method proved to be 94.07% accurate.

3.1.4   Wavelet based approximate entropy and probabilistic neural network

In [22], the authors developed an improved means for the detection of EEG patterns using wavelet based approximate entropy (ApEn) and probabilistic neural network (PNN). After applying wavelet transformation to the EEG signals, ApEn was used to facilitate epileptic detection. Out of the 60 ApEn features extracted, the best 30 were chosen, based on minimal variance within the class and maximum absolute difference between the classes. A probabilistic neural classifier was then used for classification. The dataset used was obtained from the data made available online by Dr. Ralph Andrzejak of the Epilepsy Centre at the University of Bonn, Germany. Overall accuracy as high as 100% was achieved.

Alison A. Dingle at al. [23] introduced an innovative PC-based system to enable the detection of epileptic activity. Relevant parameters of the brain waves were calculated, thresholded, and compared with the features of background activity, to determine candidate epileptiform

transients. These parameters included duration, amplitude and sharpness. The background features included amplitude, slope, duration and rhythmicity. The expert system then determined if the epileptiform events were focal or non-focal, and definite or probable. From the data tested, an average of 58% of epileptiform events were detected as definite (100% selectivity), with 0 false detections; whereas, an average of 80% were reported either definite or probable, with 9 false detections per hour.

[24] presented a new methodology centred around the time-frequency (t-f) analysis of EEG segments to help classify epileptic seizures. It made use of Short-Time Fourier Transform (STFT) and a few t-f distributions (TFDs) in order to calculate the power spectrum density (PSD) of each segment. This PSD was then used to extract several features, where each feature represented the signal's fractional energy in a specific frequency band and time window. The complete energy of the signal was also included as an additional feature, after which the entire feature set was fed to an artificial neural network (ANN). An accuracy of 89% was observed when tested on the datasets Z, O, F, N and S.

In [25], the authors evaluated epilepsy detection techniques using empirical mode decomposition of EEG signals. One of the seizure detectors described in detail, worked with various time and frequency features of IMFs (intrinsic mode functions). Of these, the time domain features included the Coefficient of Variation (VC), Median Absolute Deviation (MAD), Standard Deviation (STD), Mean Value (MV), Variance (VAR) and Root Mean Square Value (RMS). The frequency domain features consisted of Central, Mean and Peak Frequencies (CF, MF and PF), Standard Deviation Frequency (STDF), First and Third Quartile Frequencies (Q1F, Q3F), Interquartile Range (IR), 95% cumulated energy Frequency (MAXF), Asymmetry Coefficient (AC) and Kurtosis Coefficient (KC). Lambda of Wilks (WL) Criterion was then applied for final feature selection, followed by a linear discriminant analysis (LDA) for classification. A sensitivity of 69.4% and a specificity of 69.2% was achieved, when tested on the Freiburg Database. Table 3.1 shows a brief comparison of the work previously done in context of epileptic seizure detection. The table covers the work done on different databases including the one used in this research i.e. CHB-MIT Scalp database.

S/No.	Author(s)	Technique(s)	Features	Dataset	Performance Evaluation
1	A. F. Rabbi et al. [1]	Fuzzy rule-based algorithm	Average amplitude, Rhythmicity, Dominant frequency, Entropy	Freiburg Seizure Prediction EEG (FSPEEG)	Accuracy = 95.8%
2	Sharanreddy et al. [2]	MWT, FFNN, IApE		CHB-MIT Scalp EEG	Accuracy ~= 90%
3	Yusuf U Khan et al. [3]		Skewness, Kurtosis, NCOV	CHB-MIT Scalp EEG	Mean false detection rate = 1.1 false detections/hr Sensitivity = 100%
4	Ali Shoeb et al. [4]	Application of Machine Learning	Rhythmic activity, Channel identity, ECG	CHB-MIT Scalp EEG	Median false detection rate = 2 false detections/2 hrs
5	Sachee et al. [5]	Implementation of Probabilistic Neural Network using Approximate Entropy	Approximate Energy (ApEn)	NIMHANS, Bangalore Database	Accuracy = 90 (for generalized epilepsy, only)
6	Forrest Sheng Bao et al. [6]	Automated Epilepsy Diagnosis Using Interictal Scalp EEG and Probabilistic Neural Networks (PNN)	Relative Intensity Ratio (RIR), Petrosian Fractal Dimension (PFD), Higuchi Fractal dimension (HFD), Hjorth Parameters	Local database	Accuracy = 94.07%
7	A.S. MuthananthaMurugavel et al. [7]	Wavelet Domain Approximate Entropy-Based Epileptic Seizure Detection	Approximate entropy (ApEn) features		Overall accuracy (as high as) =~ 100%

8	Alison A. Dingle at al. [8]	A Multistage System to detect epileptiform activity in the EEG	Wave features: Duration, Amplitude, Sharpness Background Features: Amplitude, Slope, Duration, Rhythmicity		Selectivity = 100%, False detections = 0 (for 58% of the events detected as definite) False detections per hour = 9 (for 80% of the events reported as definite or probable)	
9	Alexandros T. Tzallaset al.[9]	Time-Frequency Analysis	Power Spectrum Density (PSD)		Accuracy = 89% (for datasets Z, O, N, F, S)	
10	Lorena Orosco et al. [10]	Epileptic Seizures Detection Based on Empirical Mode Decomposition of EEG Signals	Coefficient of Variation (VC), Median Absolute Deviation (MAD), Standard Deviation (STD), Mean Value (MV), Variance (VAR), Root Mean Square Value (RMS), Central, Mean and Peak Frequencies (CF, MF and PF), Standard Deviation Frequency (STDF), First and Third Quartile Frequencies (Q1F, Q3F), Interquartile Range (IR),	Freiburg Database	Sensitivity = 69.4% Specificity = 69.2%	
11	A. Subasi et al. [11]	Lifting-based Discrete Wavelet Transform (LBDWT), Multilayer Perceptron Neural Network (MLPNN), Logistic Regression (LR)			LR	Specificity = 90.3 Sensitivity = 89.2
					MLPNN (Backprop)	Specificity = 91.4 Sensitivity = 91.6
					MLPNN (L-M)	Specificity = 92.3 Sensitivity = 92.8

Table 3.1 Comparison of Different Algorithm

3.2 Summary

This chapter presented a brief description of the ongoing research in the field related to this study. Research work of various authors have been presented in this chapter using various techniques. Some of these include fuzzy based algorithms, multi-wavelet transform, and probabilistic neural networks and wavelet based entropy. In the end, a comparison of the previously implemented techniques along with the datasets on which the techniques were tested are presented.

Chapter 4: EEG Analysis

The process of epilepsy diagnosis is broadly broken down into three categories: pre-processing, feature extraction and classification. This chapter covers all the implementation part of this research. The figure 4.1 shows the system level diagram of the proposed automated system. It is comprised of EEG electrodes which capture the EEG signals, the acquisition box and front end.
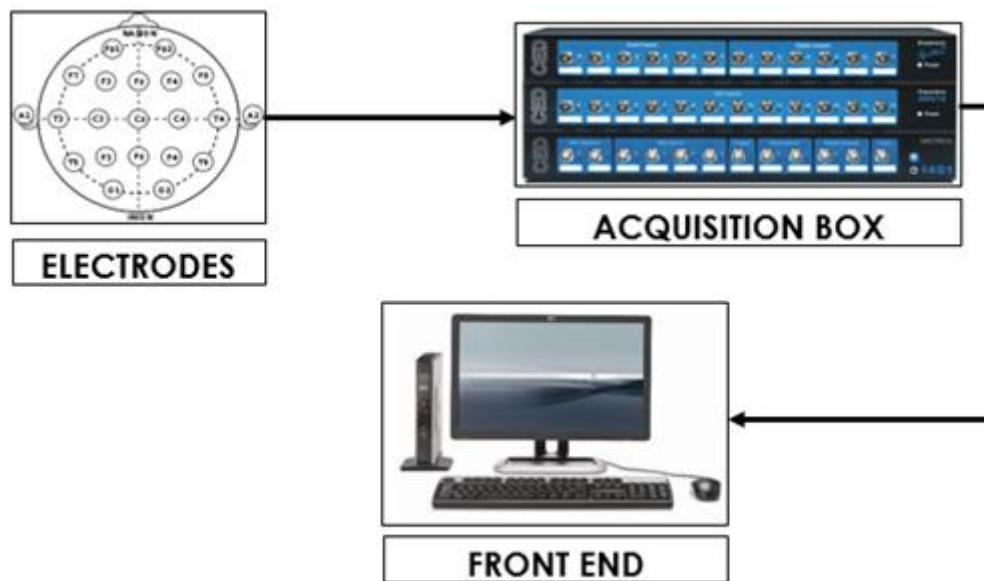
**Figure 4.1 System level Diagram**

The proposed system consists of a 4 – step process, initially the pre – processing is accomplished in which data from the CHB – MIT Scalp EEG database is organized in a manner which suits the feature extraction process. In feature extraction, a variety of temporal and spectral domain features are applied to the data set. *Ranksum* tests (Ansari – Bradly and the Wilcoxon test) have been applied in the following process to determine the accuracy obtained by the feature computed in the previous step. The best features, as determined by the techniques mentioned previously, are further tested for accuracy by using a hybrid classifier. The final Step of Classification uses the system evolved from the previous processes and helps classify any stream of data given to it. Figure 4.2 shows detailed flow diagram of our proposed method.
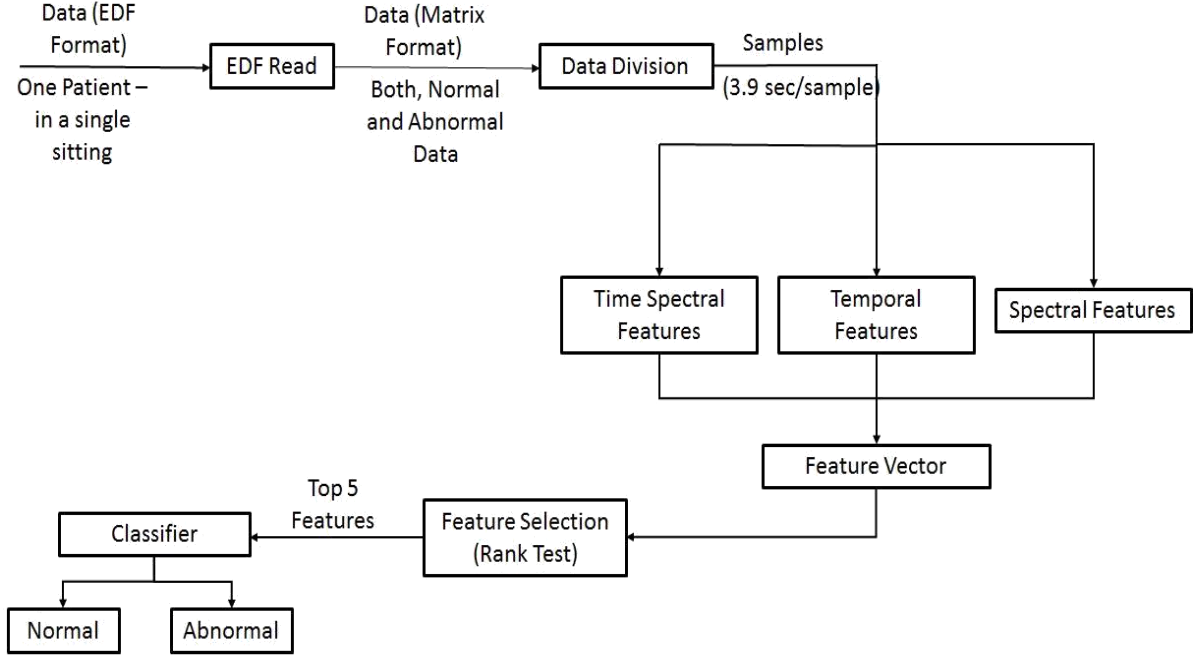
Figure 4.2 Structure of proposed method

4.1 Pre-Processing

The data, from the CHB − MIT Scalp EEG database is present in the European Data Format (EDF). It is a format commonly used to represent physical and biological signals where the signals are acquired through multi-channel observers. Furthermore, for a singular sitting, the data is grabbed for lengthy time durations (more than 60 minutes in most of the cases).

The already de − noised data is then divided into chunks, one thousand columns long, meant to facilitate the future manipulations. This is done by reviewing the patient summary files that contains information regarding the events that occur during that particular EEG sitting of the respective patient. The chunk containing information regarding the occurrence of a seizure is classified as an abnormal sample whereas one without a seizure is considered normal.

Farlang et al [23] proposed the use of surrogate channels which are gained by extending the pre-processing block with spatial filters. A large laplacian filter is used as a pre-filter which serves as a high pass filter which enhances the localized activity while suppresses the diffusion activity. In this research all the channels are converted into a surrogate channel using a large laplacian filter. If the recordings at position *(i,j)* of the skull are shown as *V(i,j)*, then the laplacian at *(i,j) is as* shown in equation 4.1.

$$((V(i,j) - 2V(i-1,j) + V(i-2,j))^2 + (V(i,j) - 2V(i,j-1) + V(i,j-2))^2 \qquad (4.1)$$

Figure 4.3 and 4.4 show the surrogate channel for EEG with seizure and without seizure respectively. It is clear from the figures that there is clear visual difference between these two waveforms. The surrogate channel with seizure have more prominent variation as compared to the waveform without seizure.
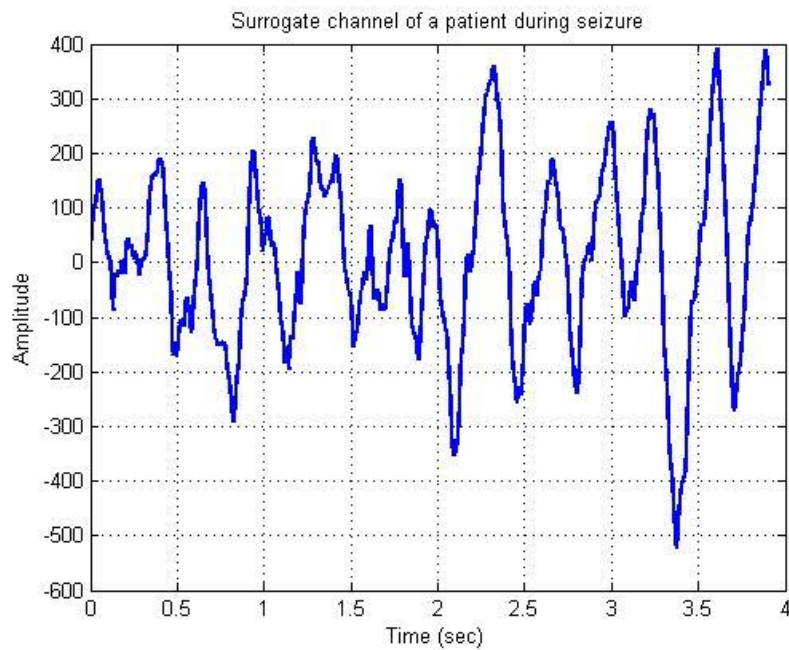
Figure 4.3 surrogate channel of a patient with seizure
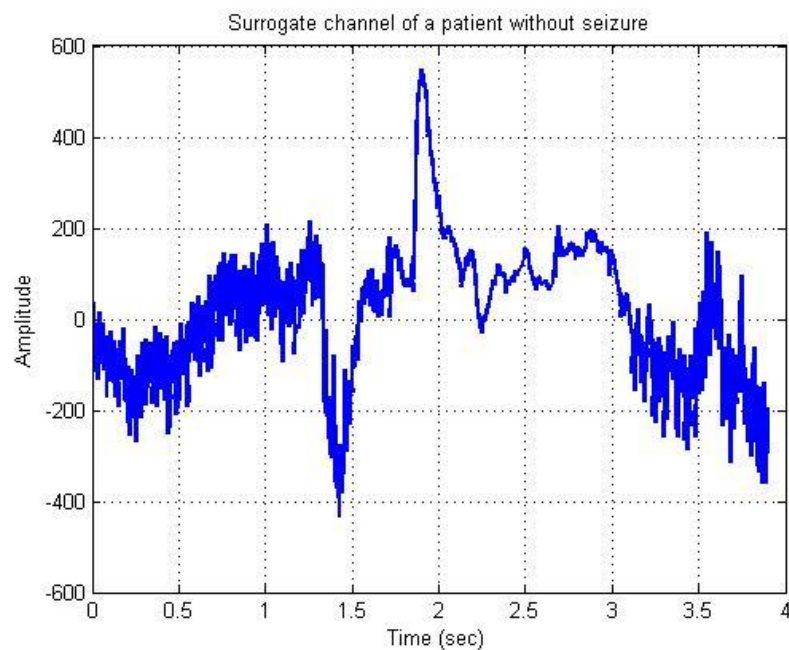
Figure 4.4 surrogate channel of a seizure without seizure

4.2 Feature Extraction

Once the pre-processing is complete, the next step is feature extraction. In this step, a number of temporal, spectral and time-spectral features are computed out of the pre-processed data.

4.2.1 Time – Domain Features

Entropy (*f1*) [7] is generally referred to as the measure of uncertainty in a given random variable. The entropy of a thousand samples is collected from a specific portion of the brain, received via a distinct channel. The type of entropy applied to the data set is 'Shannon' (non-normalized), which is the summation of the multiplication of each sample squared with the log of the squared sample as shown in equation 4.2.

$$E(s) = -\sum_i s_i^2 \log(s_i^2)$$
(4.2)

Mean (*f2*) is the measure of central tendency of a given probability distribution. The mean of a set of data points is calculated by multiplying each point with its probability and then summing them together. The resulting value is then divided by the range of data points used for a single-value analysis. This is depicted in equation 4.3

$$M(s) = \frac{1}{n}\sum_i s_i P(s_i)$$
(4.3)

Harmonic Mean (*f3*) [8] refers to a specific kind of averaging technique. It is calculated by summing the reciprocals of individual signals and then, computing it's reciprocal as shown in equation 4.4. The resulting value is then multiplied by the total number of data points, used for calculation.

$$HM(s) = \frac{n}{\sum_i \frac{1}{s_i}}$$
(4.4)

Range (*f4*) is the difference between the maximum and the minimum values in a given data set as seen in equation 4.5. Range for the data, acquired from each channel, is calculated in order to understand the domain of the acquired values.

$$R(s) = \max(s_i) - \min(s_i)$$
(4.5)

Inter Quartile Range (*f5*) Inter quartile range is broadly referred to as the measure of statistical dispersion of the data. It is the difference between the 25th and the 75th quartile of the spread of the information, as collected from a single EEG channel. Equation 4.6 describes this feature.

$$IQR(s) = Q_3(s) - Q_1(s) \qquad (4.6)$$

Where $Q_3(s) = (\sigma(s) * z_3) + M(s)$, $Q_1(s) = (\sigma(s) * z_1) + M(s)$, $\sigma(s)$ = Standard Deviation and M(s) = Mean.

Mean Absolute Deviation (*f6*) Mean Absolute Deviation is referred to as the mean of the absolute deviations about the mean of a set of data. It calculates the average distance from the mean of a thousand data points obtained from a channel as shown in equation 4.7

$$MAD(s) = E|X(s) - \mu(s)| \qquad (4.7)$$

Moment (*f7*) is in general, a quantitative measure of the shape of points in a given data set. It determines the central moment of the signals under study.

Skewness (*f8*) [11] is the third central moment which determines the lop-sidedness of a set of data. A symmetric distribution would have a skewness of zero, whereas a distribution which is skewed to either the left or the right side with respect to the centre would have a negative or a positive skewness value, respectively. Equation 4.8 gives a mathematical interpretation of this feature.

$$S(s) = \frac{E\left(\left(X(s) - \mu(s)\right)^3\right)}{\sigma^3} \qquad (4.8)$$

Kurtosis (*f9*) Kurtosis being the fourth central moment, is the measure of the 'peak width' of a normal distribution. It determines if the data, under study, is tall and thin or short and squat when compared to the normal distribution. Equation 4.9 governs this feature.

$$K(s) = \frac{E\left(\left(X(s) - \mu(s)\right)^4\right)}{\sigma^4} \qquad (4.9)$$

Percentile (*f10*) indicates the value below which a given percentage of values in a group of values fall.

Gradient (*f11*) magnitude represents the slope of a set of points. It points towards the greatest rate of increase in the under study data points.

4.2.2 Frequency-Domain Features

Wavelet Transform is the type of a time-frequency transform which follows the basic idea of changes in time-extension without altering its overall shape. The wavelet decomposition vector (*f12*) and the bookkeeping vectors (*f13*) are formed by choosing the appropriate basis

functions. A multi-level one dimensional wavelet decomposition is performed using 'db1' as the shape of the wave to work upon.

Wavelet Energy calculates the energy for a 1-D wavelet packet decomposition. The percentage of energy which corresponds to the approximation (*f14*) and the percentage of energy corresponding to the details (*f15*) of the outputs of the wavelet transform feature are calculated to comprehend the energy details of the wavelet decomposition, calculated prior to this.

Pseudo spectrum refers to a set that contains the spectrum of the operators and the numbers which are almost Eigen values. The pseudo spectrum estimate (*f16*) of the input data and the normalized frequencies (*f17*), where the estimate is calculated, are found. These values are calculated by using the correlation matrix of the data under study and utilizing its Eigen vector estimates.

Fast Approximate Entropy (*f18*) calculates the amount of irregularity over the time-series of a given data set. A set of data that repeats its pattern over time is considered more predictable as opposed to a situation where it doesn't. ApEn outputs the likelihood relating the occurrence of consecutive similar patterns. A set of data with a high degree of randomness has a greater ApEn value whereas a less random data would have a comparatively lesser ApEn value.

4.3   Feature Selection

All the aforementioned features have been tested on the relevant EEG data set, results of which are seen to vary considerably from feature to feature. This variety in the obtained results makes the feature-selection procedure all the more critical, as a feature with poor outcomes hinders the much needed good sensitivity and specificity results.

To achieve the desired goal, two types of statistical tests are carried out: Wilcoxon signed - rank test and the Ansari-Bradley test. Both these tests are applicable to two-sample designs.

The Wilcoxon test measures the degree of similarity between the two samples according to the class median value. The lower the value, the better the feature under consideration since this helps us classify the abnormal and the normal samples to a greater accuracy. The Ansari-Bradley test, on the other hand, takes into account the distribution shape of the feature and calculates the degree of dispersion according to it. The same rule for the degree of accuracy stands true for this test too.

4.3.1 Wilcoxon Signed Rank Test

Wilcoxon signed rank test is a non-parametric statistical test. This test is further classified as single sample and paired sample test. Single sample test uses dingle sample and is used whenever we desire to test a hypothesis about population median. Whereas paired test uses two samples which should be paired. Each pair is chosen randomly and independently. Let us assume that N is the sample size, then there are 2N data points. Equation 4.10 governs this test.

$$W = |\Sigma_{i=1}^{N_r}[sgn\,(x_{2,i} - x_{1,i}).R_i]| \tag{4.10}$$

The absolute value of the sum of the signed ranks where x1,i and x2,i represents the measurements, Ri is the rank and W is test statistic.

Wilcoxon signed rank test neither depends on parent distribution nor on its parameters. It does not require any assumption about the shape of the distribution. Efficiency of this test is almost 95%.

4.3.2 Ansari-Bradley Test

Ansari - Bradley test, on the other hand, is a distribution free test. It is used if one distribution has different variability than the other one. We assume same median in this test. The algorithm of this test is given below.

H0 : F(t) = G(t) for all t
Assume same median θ1 = θ2
Assume that F(t)=H((t- θ1)/ŋ1)
And G(t)=H((t- θ2)/ ŋ2)
X= ŋ1Z + θ1, Y= ŋ2Z + θ2, with Z ~ H
H is continuous with median 0 ⇒ F(θ1) = G(θ2) = 1/2
 If θ1 ≠ θ2, but both are known, shift each sample:
                    X'i = Xi − θ1
Y'i = Yi − θ2. Now, we have a common median 0
γ = ŋ1/ŋ2 where γ is the ratio of scales.

Algorithm 1: Algorithm for feature selection based on Ansari Bradley Test

4.4 Classification

The determination of the final feature set helps us classify any EEG signal that serves as an input to our classifier, as normal or abnormal, depending on the information contained in it. The data is divided into two main classes:

$$C_1 = \{Abnormal\ or\ Seizure\}$$

$$And,\quad C_2 = \{Normal\ or\ non-Seizure\}$$

The classification methodology applied is supervised in which the data is randomly divided into two distinct subsets: the testing and the training sets. Initially, the classifier goes through a training phase with the help of the labelled data, which assigns a class to every sample. After that we use the testing data to understand the extent of accuracy as achieved by the classifier under consideration.

A hybrid classifier consisting of three well known classifiers serves as classifier for the proposed system. The constituent classifiers are as mentioned below.

4.4.1 k-Nearest Neighbours (kNN)

kNN is the most simple and fundamental classifier used for supervised classification [15]. It is a kind of voting based classifier which finds k-nearest samples from complete dataset based on some distance calculation between training and test samples. Let vi be a feature vector for ith node with m features ($fi1, fi2, fi3, \ldots, fim$); let n be the total number of nodes ($i = 1, 2, \ldots, n$) and let m be the total number of features ($j = 1, 2, \ldots, m$). The Euclidean distance between nodes vi and vl where ($l = 1, 2, \ldots, n$) is defined as in equation 4.11.

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \ldots + (x_{ip} - x_{lp})^2} \qquad (4.11)$$

Now, depending upon the value of k, we choose closest k samples and assign the majority class to unknown node.

4.4.2 Gaussian Mixture Model (GMM)

To implement GMM, we use a two-class Bayesian classifier using Gaussian functions. Bayes decision rule is stated before. Choose $R1$ if, $p(\text{k} \mid R1) \, P \, (R1) > p(\text{k} \mid R2) \, P \, (R2)$ otherwise choose $R2$, where $p(\text{k} \mid Ri)$ is the class conditional probability density function (pdf) also known as likelihood and $P(Ri)$ is the prior probability of class Ri which is calculated as the ratio of class Ri samples in the training set. The class conditional pdf of the feature vector for different classes is computed using multivariate Gaussian pdf as given in equation 4.12.

$$N(v|\mu, \Sigma) = \frac{1}{2\pi^{m/2}|\Sigma|^2} \exp[-\frac{1}{2}(v - \mu)\Sigma^{-1}(v - \mu)] \qquad (4.12)$$

Where v and μ are feature vector containing m number of features and mean vector containing mean of each feature, respectively. Σ is a $m \times m$ covariance matrix. In our case, $m = 7$. We model the class conditional pdf's as linear combination of weighted Gaussian functions to represent the likelihood of a GMM as in equation 4.13.

$$p(v|R_i) = \Sigma_{j=1}^{K_i} N(v|\mu_j, \Sigma_j) w_j) \tag{4.13}$$

Where K_i is the number of Gaussian mixtures used for Bayesian classification, $p(v|R_i)$ is a m-dimensional Gaussian distribution of weight w_j, and $Ri = \{R1, R2\}$ are the two classes used in proposed system. Equations (4.12) and (4.13) show the likelihoods for a single Gaussian distribution and GMM, respectively. The parameters for GMM are optimized using expectation maximization (EM) which is an iterative method and it chooses optimal parameters by finding the local maximum value of GMM distributions for training data. The EM starts with initial values of parameters (μ, Σ) and weight w for each Gaussian. In estimation step, EM computes the probability (PE) of each point for each Gaussian using equation 4.14.

$$P_E(n,j) = \frac{w_j \, N\left(v_n \,|\, \mu_j \,,\, \Sigma_j\right)}{\Sigma_{i=1}^{k} N\left(v_n \,|\, \mu_i \,,\, \Sigma_i\right) w_j} \tag{4.14}$$

Here, $P_E(n, j)$ represents the probability that nth candidate region v_n is generated from jth Gaussian. We do this for all κ Gaussians and candidate regions. The second step is the maximization of likelihood by changing the parameters. The mean, covariance matrix, and weight for jth Gaussian are updated using estimated probabilities and are given in equation 4.15, 4.16, and 4.17, respectively.

$$\mu_j = \frac{1}{\xi_j} \Sigma_{n=1}^{N_{Total}} P_E(n,j) \, v_n \tag{4.15}$$

$$\Sigma_j = \frac{1}{\xi_j} \Sigma_{n=1}^{N_{Total}} P_E(n,j) \, (v_n - \mu_j)(v_n - \mu_j)^T \tag{4.16}$$

$$w_j = \frac{\xi_j}{N_{Total}} \tag{4.17}$$

4.4.3 Support Vector Machine (SVM)

SVM is the third classifier proposed in this framework to detect epileptic seizure. The SVM classifier separates the data belonging to one class from the data belonging to another class by specifying a hyperplane keeping the separation between the classes maximum. The nature of the data in hand led us to features which make a non-linear hyperplane generated by the SVM with the kernel function based on radial basis function (RBF). The least square SVM technique has been implemented using the LS-SVM toolbox [18]. In LS-SVM, the multiclass solution is found by solving a system of linear equations instead of original quadratic programming.

4.4.4 Hybrid Classifier (HC)

The hybrid classifier proposed in this study combines the kNN, GMM and SVM classifiers. It assigns a weight to the results of all these individual classifiers and uses this weighted probabilistic ensemble for further processing. The classification is carried out based on the measure of evidence produced by the individual classifiers as shown in equation 4.18.

$$class(v) = arg.\,max_{\forall class_i}(\Sigma_{i=1}^{c}a_k * P_{C_k}(y = class_i|v))\qquad(4.18)$$

Where $P_{C_k}(y = class_i|v)$ is the probability of class$i$ given a sample node using classifier k and ak is the weight associated with the probabilistic prediction of class C_k. Figure 4.5 shows the proposed ensemble framework for hybrid classifier.
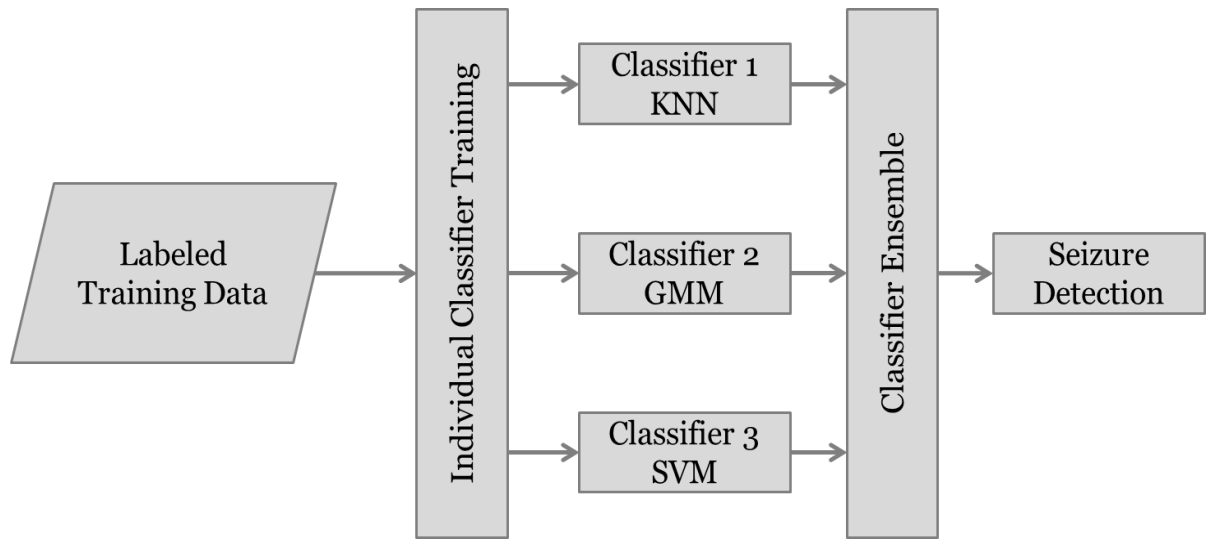

Figure 4.5 Block diagram of Hybrid Classifier

4.4.5 Learning Optimized Weights Using Genetic Algorithm

The proposed framework carrying an ensemble consists of a feature vector $ak = \{ak$NN, aGMM, aSVM$\}$. These weights in this weight vector are optimized using genetic algorithm. The modelling of weights is composed of two phases. In the first phase, the separation of confused samples from the rest of the samples present in the training dataset is carried out. Confused samples are those samples which are classified to different classes when classified by the individual classifiers. Since these are the only samples for which the results of the hybrid classifier is pressed down, we intend to use these samples only to optimize our proposed framework. As it takes less time to process only the confused samples, in second phase, weights are optimized for the confused samples determined in the previous phase. This optimization is carried out using the genetic algorithm. The parameters involved in the genetic algorithm such

as the size of population, rules for crossover, the definition of population and mutation are described below.

(i) Population:

Each chromosome in our framework consists of a weight vector. This weight vector has three members which are weights for each classifier. The values of these weights are kept as normalized for ease in further processing.

(ii) Population Size:

The size of the population in this study is 20 normalized weight vectors. Out of these, 16 are generated randomly and the remaining four are generated keeping maximum probability (confidence) for individual classifiers and equal probability for all the classifiers as [1, 0, 0], [0, 1, 0], [0, 0, 1], and [0.33, 0.33, 0.33].

(iii) Crossover:

Crossover can be a single point or multiple point crossover. In this study, we have used single point uniform crossover. Crossover point is kept just after the first weight element which means that two selected chromosomes exchange their weights for GMM and SVM classifiers keeping the same weight for kNN classifier. The value presented by an objective function selects the chromosomes. The worst 10 out of the total 20 chromosomes are selected for crossover.

(iv) Mutation:

The probability of mutation in this study is kept 0%. It shows that there is no change inserted in offspring after crossover.

(v) Objective Function:

The accuracy achieved by an individual classifier with respect to a specific weight vector has been determined as the objective function. As better accuracy is always desired, we want to maximize the accuracy, in other words, the objective function.

The iterative learning method is applied until there is no improvement observed in classification accuracy given in the equation for ten consecutive iterations or the algorithm reaches to maximum iteration which is kept to 100 in this study.

4.5   Summary

A detailed explanation of the proposed method is given in this chapter. The input EEG is passed through pre-processing phase initially where a surrogate channel is generated out of the data available for all the channels. Then features of time, frequency and time-spectral domain are extracted out of the data. Then the best features are selected using Ansari Bradley and Wilcoxon rank tests. The best features are used to make a hybrid feature set. Then the classification is carried out with kNN, SVM and GMM classifiers separately. These classifiers are optimized over the available training data and then the optimized classifier (of all the three types) are fed to the proposed hybrid classifier. The hybrid classifier is also tuned for the best parameters using the genetic algorithm. The hybrid classifiers, finally, classifies the test sample as either normal or abnormal.

Chapter 5:   Experimental Results

This chapter contains information regarding the material used for the evaluation of the proposed system. It also contains results as obtained at different stages of the work. We have also compared our results with the state of art techniques.

5.1   Dataset

Two datasets were used for the purpose of evaluation. One is of the CHB- MIT scalp EEG database. This database is collected at children's hospital in Boston. Recordings are collected from 5 males (age 3-22) and 17 females (age 1.5-19). These recordings are grouped into 23 cases. Each case contains 9 to 42 continuous .edf files from a single subject. There are gaps in the recordings due to hardware problems. In most cases gaps are usually 10 seconds or less. Occurrence of longer gaps is also seen but occasionally. .The recordings of the EEG files are of exact one hour in time.

Sampling rate of all the signals used is 256 samples per second with 16 bit resolution. 10-20 electrode placement system is used for the recordings of these EEG signals. The proposed method is applied to a total of 12 patients.

5.2   Performance Measures

Validity of the proposed system is determined using different performance measures like sensitivity (true positive rate), specificity (true negative rate) and overall accuracy. These parameters were calculated using equations 5.1-5.3 respectively.

$$\text{Sensitivity= TP/(TP+FN)} \tag{5.1}$$

$$\text{Specificity= TN/(TN+FP)} \tag{5.2}$$

$$\text{Accuracy= (TP+TN)/(TP+TN+FP+FN)} \tag{5.3}$$

Where,

- TP are true positives, meaning segments with seizures which are correctly classified.

- TN are true negatives, meaning normal segments which are correctly classified.

- FP are false positives, meaning normal segments are wrongly classified as seizure.

- FN are false negatives, meaning segments with seizures are wrongly classified as normal.

5.3   Results

Table 5.1 shows the results of rank-sum tests applied to the given data.

All the aforementioned features have been tested on the relevant EEG data set, results of which are seen to vary considerably from feature to feature. This variety in the obtained results makes the feature-selection procedure all the more critical, as a feature with poor outcomes hinders the much needed good sensitivity and specificity results.

| FEATURES	ANSARI – BRADLEY TEST		FEATURES	WILCOXON TEST	
	p-value	Score		p-value	score
f18	$< 10^{-6}$	5.47	f7	$< 10^{-6}$	5.02
f16	$< 10^{-6}$	6.67	f3	$< 10^{-6}$	5.86
f12	$< 10^{-6}$	9.27	f9	$< 10^{-6}$	6.13
f1	$< 10^{-6}$	9.4	f18	$< 10^{-6}$	7.05
f3	$< 10^{-6}$	10.04	f14	$< 10^{-6}$	11.80
f10	$< 10^{-6}$	10.38	f15	$< 10^{-6}$	11.80
f11	$< 10^{-6}$	12.06	f4	$< 10^{-6}$	22.58
f6	$< 10^{-6}$	12.24	f16	$< 10^{-6}$	23.02
f5	$< 10^{-6}$	15.40	f1	$< 10^{-6}$	24.9
f7	$< 10^{-6}$	18.56	f6	$< 10^{-6}$	26.21
f2	$2.61 * e^{-5}$	3.94	f5	$< 10^{-6}$	27.15
f3	$8.04 * e^{-5}$	4.2	f8	$4.72 * e^{-5}$	4.07
f8	$8.64 * e^{-4}$	3.33	f10	0.011	2.54
f9	0.0238	2.26	f12	0.025	2.24
f14	0.62	0.50	f2	0.67	0.43
f15	0.62	0.50	f11	0.8747	0.16
f17	1	$2.09 * e^{-6}$	f13	1	$2.15 * e^{-6}$
f13	1	$2.1 * e^{-6}$	f17	1	$2.22 * e^{-6}$

Table 5.1 Rank-sum Tests Results

To achieve the desired goal, two types of statistical tests are carried out: Wilcoxon signed - rank test and the Ansari-Bradley test. Both these tests are applicable to two-sample designs.

The Wilcoxon test measures the degree of similarity between the two samples according to the class median value. The lower the value, the better the feature under consideration since this helps us classify the abnormal and the normal samples to a greater accuracy. According to the results shown in the table above, features 7, 3, 9, 18, 14, 15, 4, 16, 1, 6 and 5 are able to distinguish between the two classes with the best accuracy.
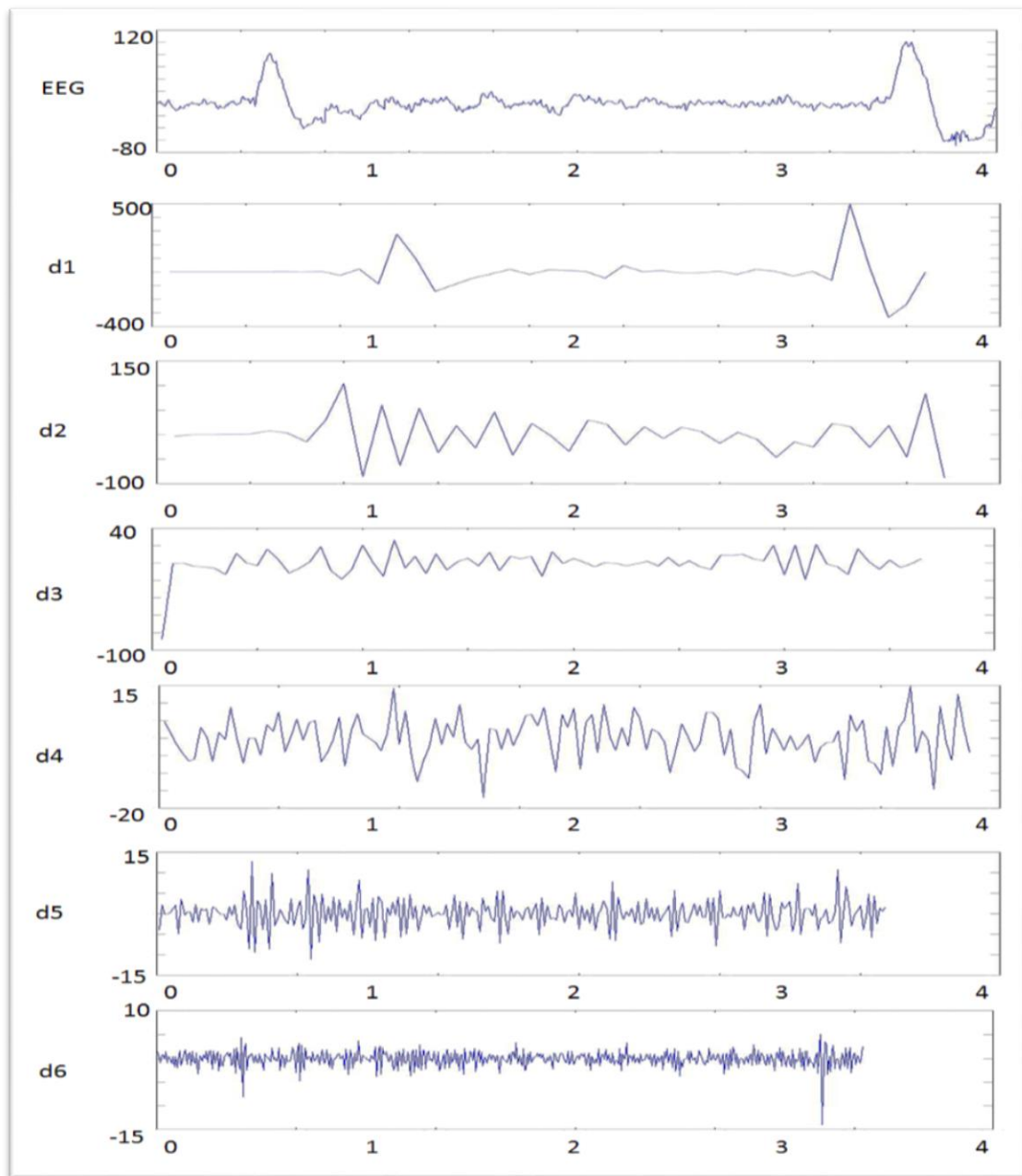
**Figure 5.1 Normal patients (wavelet decomposition)**
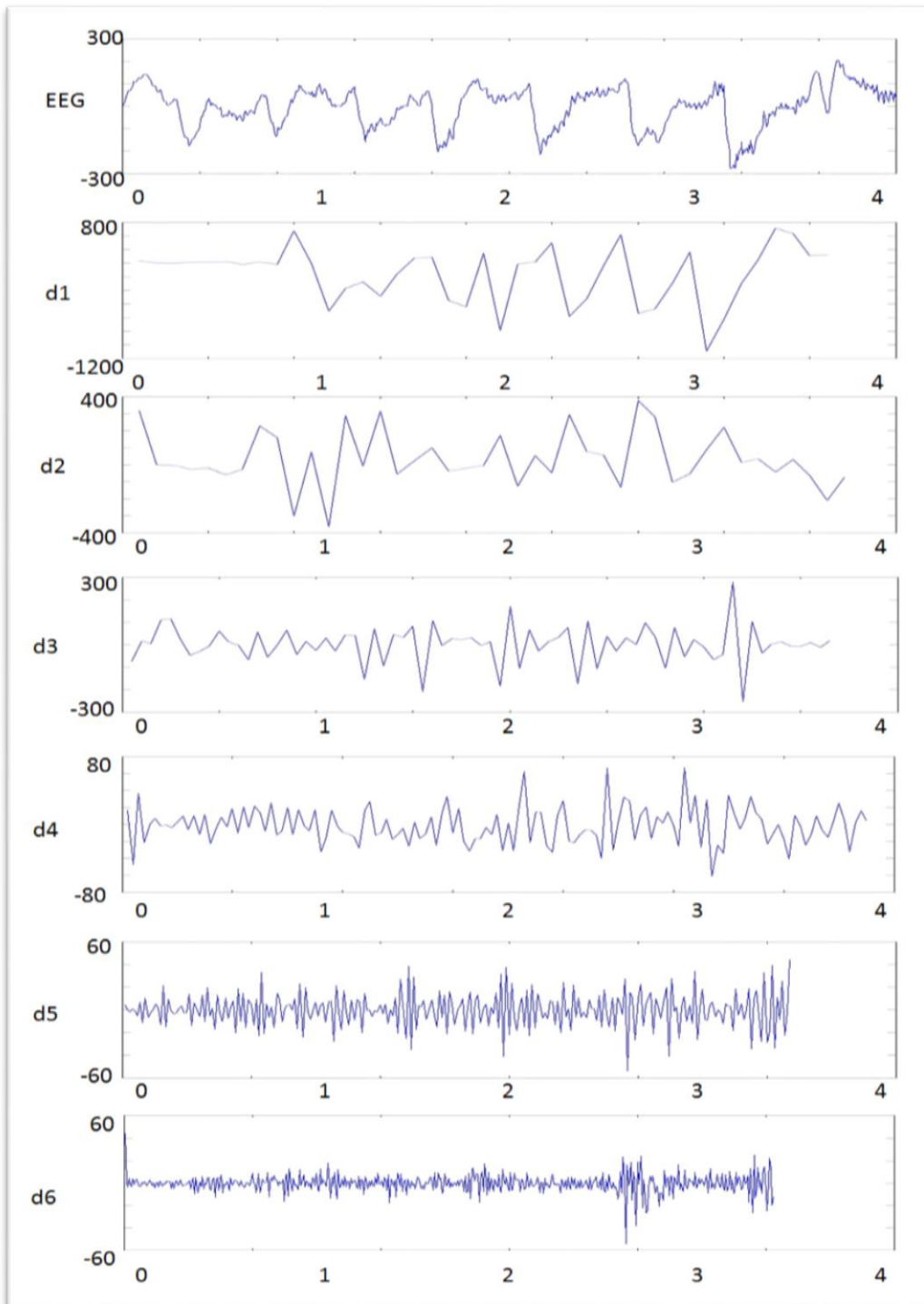
Figure 5.2 Abnormal patients (wavelet decomposition)

The Ansari-Bradley test, on the other hand, takes into account the distribution shape of the feature and calculates the degree of dispersion according to it. The same rule for the degree of accuracy stands true for this test too. The best features obtained through this test are: 18, 16, 12, 1, 3, 10, 11, 6, 5 and 7. Figures 5.1 and 5.2 show figurative results of wavelet decomposition

for normal and abnormal EEG. The rhythmicity and high variation at different level of decomposition are very clear in EEG data of abnormal patient and it is absent in normal case.

The evaluation of proposed hybrid classifier has been done with existing state of the art classifiers i.e. kNN, GMM and SVM. In order to perform detailed evaluation, all three classifiers have been modelled and tested using different settings. Figure 5.3 shows the results of the kNN classifier where the values of k is varied from 1-9. It is evident from the figure 5.3 that the kNN classifier yields the best accuracy for 3 nearest neighbours.
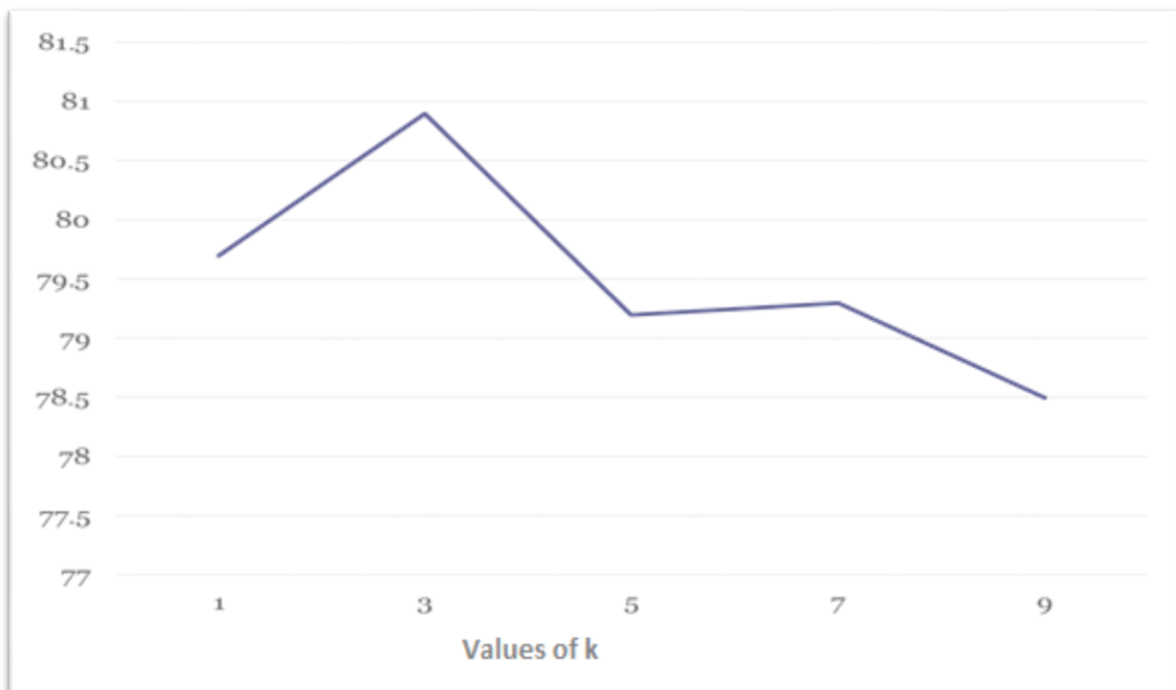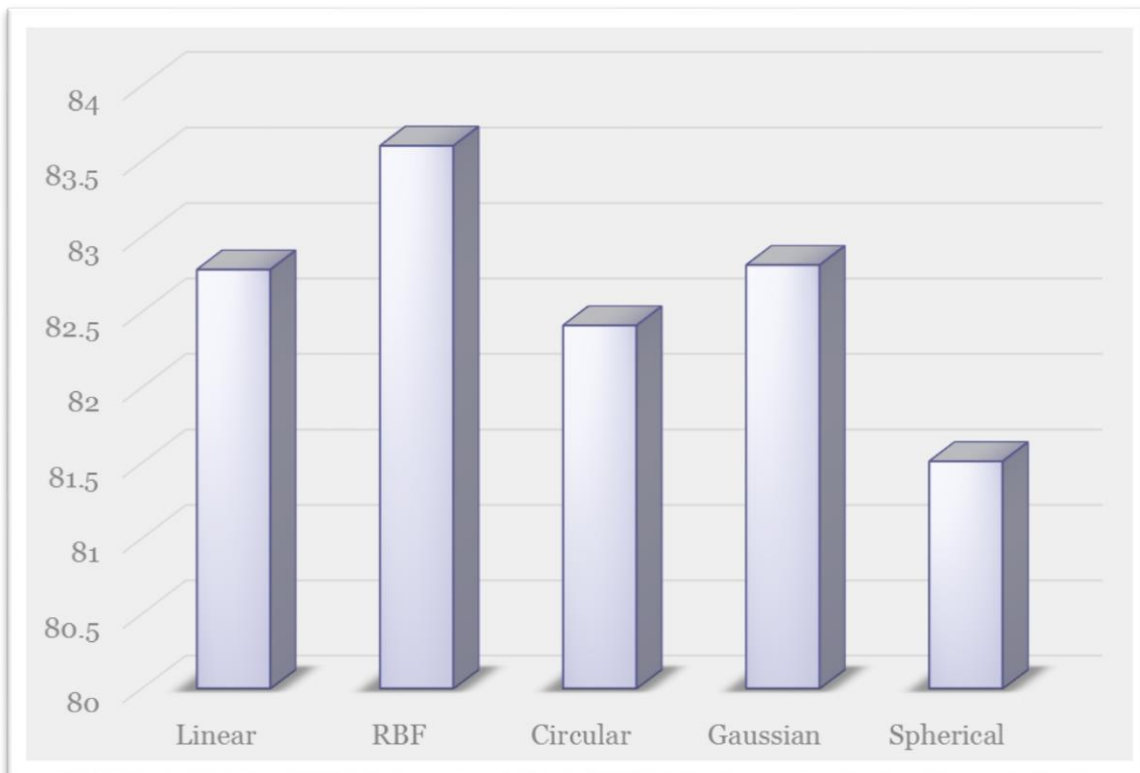

Figure 5.3 Results of kNN classifier with different values of k

Figure 5.4 depicts the results of Gaussian Mixture Models where the number of Gaussian kernels are varied over a range of 3-18. Initially the classifier is seen with improved accuracy but when the number of kernels approach 12, the improvement in accuracy is reduced. So we chose to use 18 kernels for the GMM classifier as it yielded the best accurate results.

After extracting best accuracy for GMM classifier, we opted for the Support Vector Machines for the classification. The SVM classifier was tested for the available data with different kernels. The tested kernels include the Linear, Radial Basis Function, Circular, Gaussian and Spherical kernels. Out of all the tested kernels, the Radial Basis Function gave the best accuracy as seen in the figure 5.5
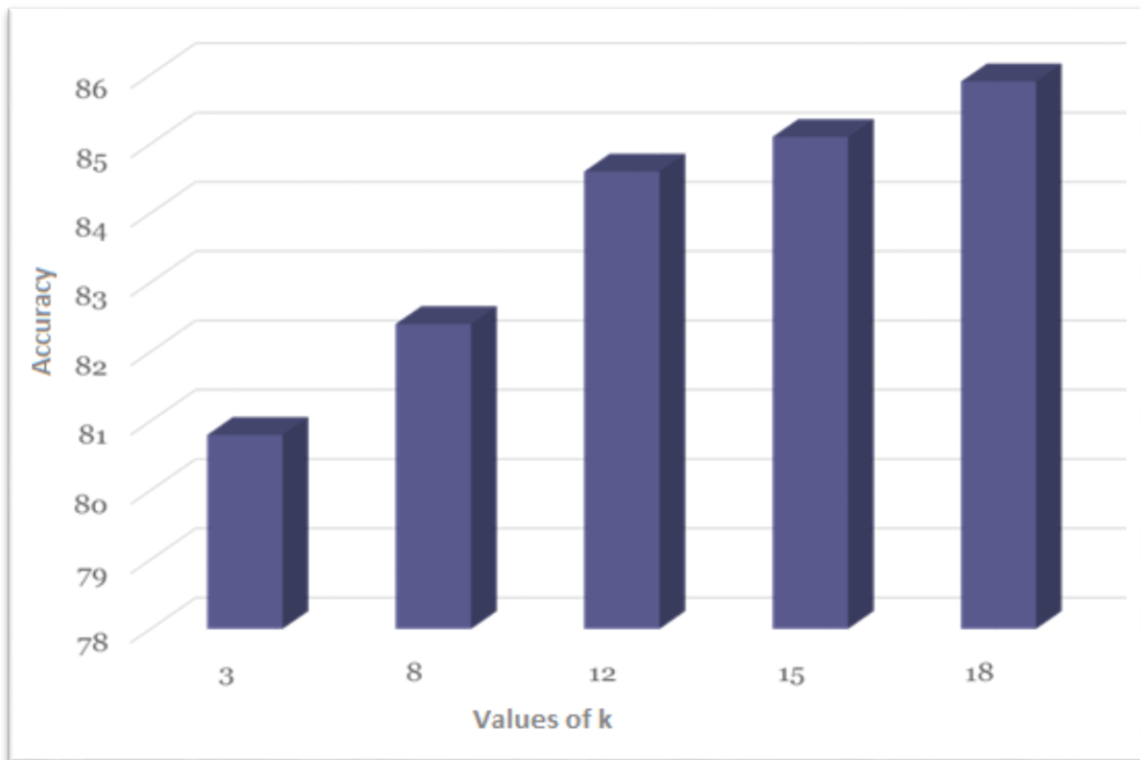
Figure 5.4 Effect of number of Gaussians (k) on overall accuracy

**Figure 5.5 Results of SVM classifier with different Kernels**

Tables 5.2 shows results of existing classifiers for each feature. The classifiers have been used against their best settings.

FEATURES	kNN (%)	GMM (%)	SVM (%)
Entropy	51.47	62.29	53.82
Mean	50.01	53.49	59.30
Range	60.23	68.11	63.95
Inter Quartile Range	84.57	87.87	88.70
Harmonic Mean	54.54	47.84	61.96
Mean Absolute Deviation	81.76	84.72	87.04
Moment	56.87	52.99	57.31
Skewness	57.23	57.64	57.48
Kurtosis	62.42	59.63	64.62
Percentile	65.98	66.11	67.77
Gradient	56.34	57.97	53.65
Wavelet Transform	62.41	59.80	64.78
Wavelet Energy	42.56	46.35	53.65
Pseudo Spectrum	58.88	64.12	54.98
Fast Approximate Entropy	52.47	64.62	54.32

Table 5.2 Results obtained by different classifiers

Figures 5.6 shows the results for hybrid classifiers with different cross validation techniques and by varying the feature vector size. It shows the averaged results for different value of k in k-fold cross validation. Vertical bars shows the standard deviation against each value of k. The best results have been achieved for 6 fold cross validation. Figure 5.7 shows the effect of feature selection on overall accuracy. The accuracies has been calculated by choosing top features based on rank tests. The best results have been achieved for top 7 features. Next all results have been carried out using 6 fold cross validation and using top 7 features.
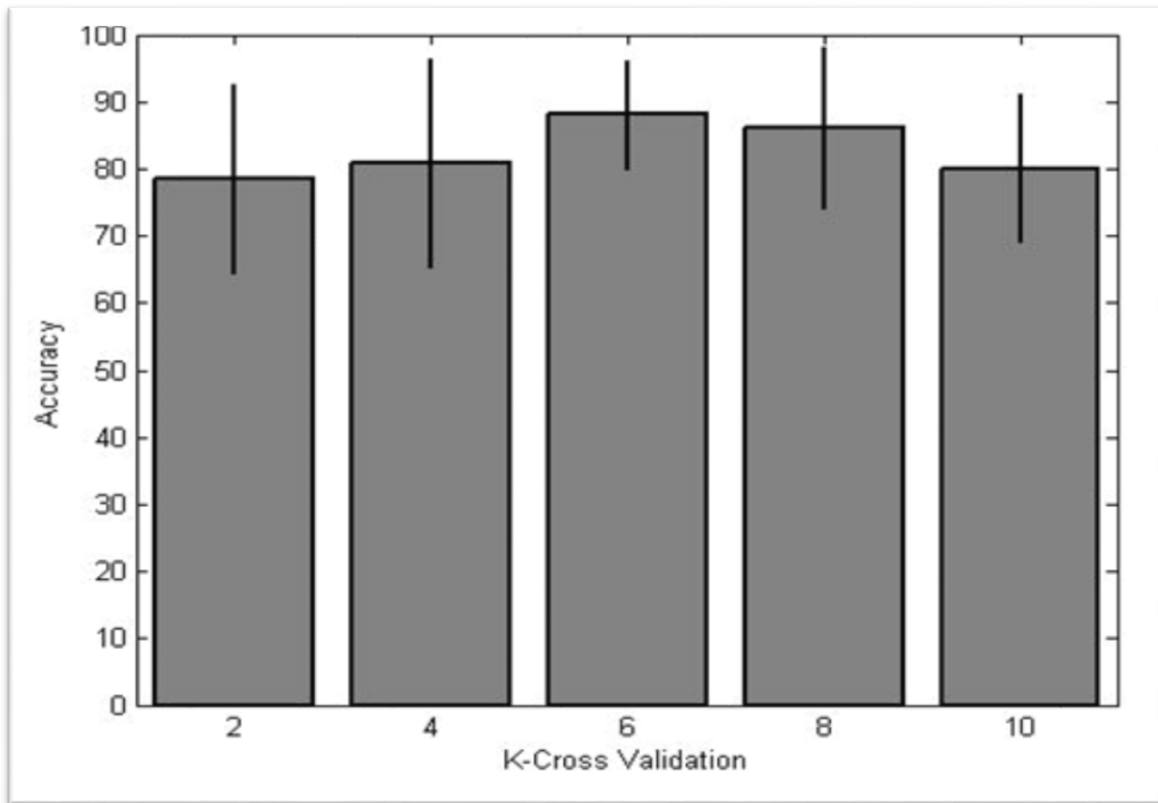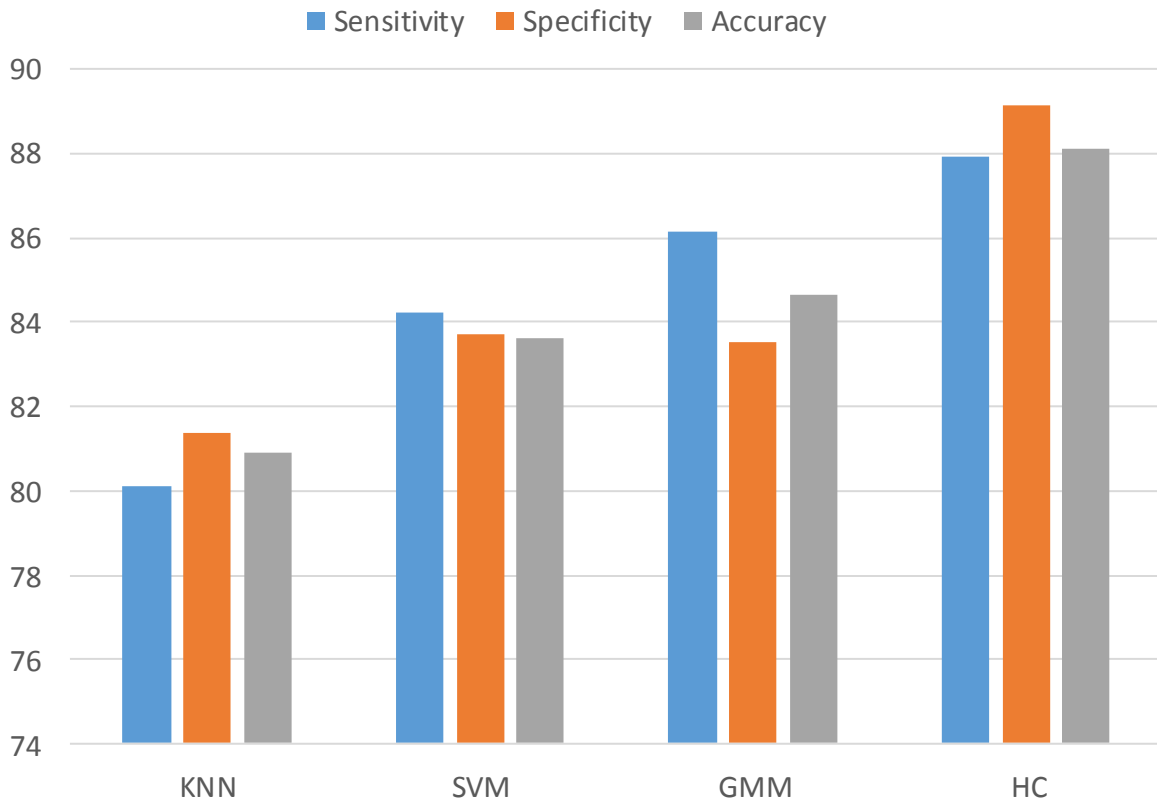
Figure 5.6 Results of hybrid classifier for different cross validations
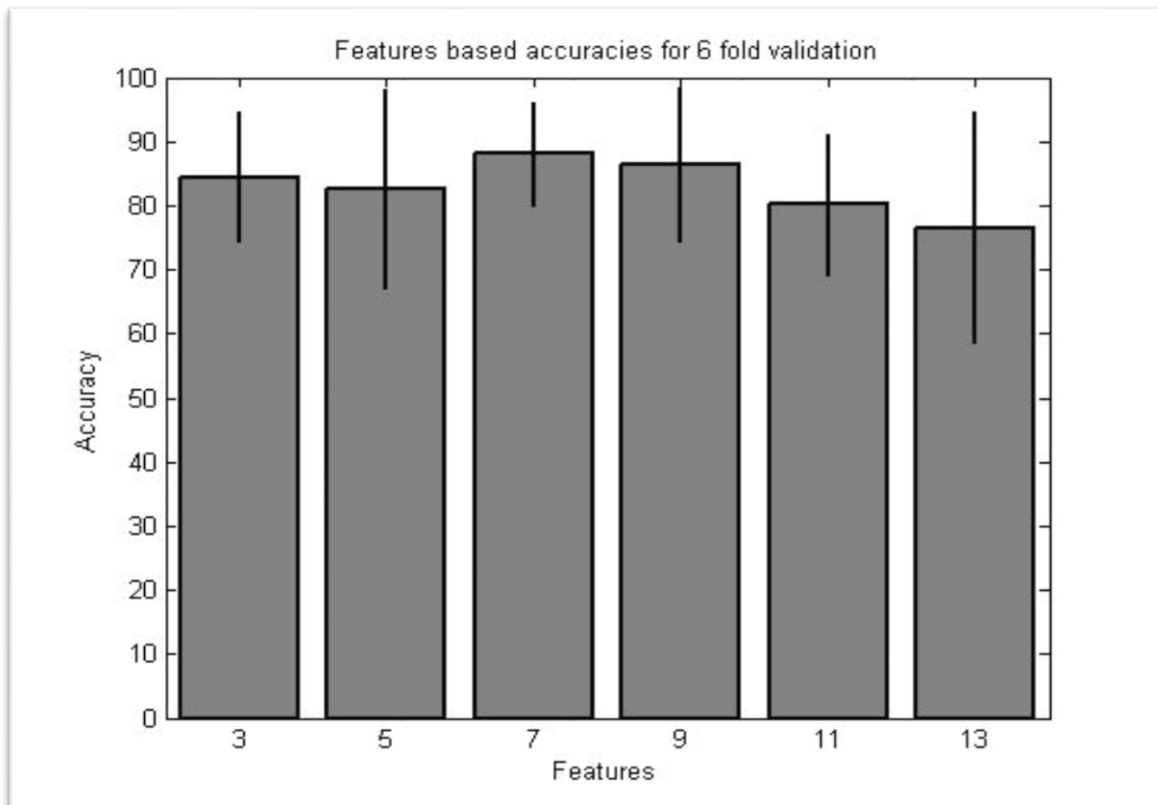
Figure 5.7 Effect of feature selection on accuracy using HC

The comparison of proposed HC with kNN, SVM and GMM is shown in figure-5.8. We have included the best results which kNN, SVM and GMM classifiers are giving for their different parameters. It is clear from figure that HC performs better than these classifiers.

Figure 5.8 Comparison of Proposed HC with kNN, GMM and SVM

Table 5.3 shows results of Hybrid classifier applied on data. Detailed results with respect to sensitivity, specificity and accuracy are presented in this table. This table shows the averaged results of all 6 folds for each patient data individually and also shows the averaged results for proposed hybrid classifier on whole data.

Patient	Sensitivity	Specificity	Accuracy
1	94.9	96.5	95.3
2	94.9	94.4	94.4
3	89.5	93.4	90.9
4	82.3	82.5	82.1

5	99.9	99.4	99.6
6	87.8	91.1	87.5
7	94.1	90.3	91.4
8	81.3	85.3	81.5
9	94.7	96.3	95.7
10	79.3	78.4	79.9
11	79.8	82.1	81.2
12	76.7	80.4	77.7
Averaged	87.94	89.17	88.10

Table 5.3 Performance Evaluation Results

5.4 Summary

This chapter presents the results and observations as seen in this research work. The dataset is explained at the beginning of this chapter followed by the performance metrics. The results section includes the results of the rank tests which help us in choosing best features. The results of the individual classifiers is presented after rank tests. The classifier with optimized parameters giving best accuracy is passed to the hybrid classifier. The hybrid classifier is also tested against various cross validations, 6 fold came out to be the best. Similarly best number of features to be used for classification using hybrid classifier came out to be 7. In the end detailed results with respect to the performance metrics for hybrid classifiers have been presented revealing accuracy of 88.10%.

Chapter 6: Conclusion and Future Work

Epileptic seizures are affecting a large amount of population in the world. The normal procedure includes the acquisition of EEG which is then observed by an expert neurologist who terms it as normal or patient. This procedure is both time consuming and requires expert neurologists as well which is not available in abundance in our country. To help the patients, an automated system to replace this entire procedure is need of the time. It has been determined that Digital Signal Processing techniques are very crucial to the automation of seizure detection. Research institutes have worked exhaustively to come up with methodologies in this domain. The automated system proposed in this research work presents a novel technique that focuses on automating seizure detection, under-going four distinct steps: Pre - processing, feature extraction, feature selection and classification. The pre-processing generates an equivalent signal out of the input signal. The feature extraction part extracts several features out of the equivalent signal. These features are from time, frequency and time-spectral domain. The best features are selected from the available set of features and passed over to classification block. The classification of the signal is governed by a hybrid classifier, which carries an essence of three popular classifiers in the domain, optimized for the best results. The proposed system has achieved average values of sensitivity, specificity and accuracy as 87.94%, 89.17% and 88.10% respectively. The results have shown the validity of proposed system as it has detected a larger number of seizure samples correctly.

For now, I am only working on the processing of the data to detect epileptic seizure but in the long run, a cost effective portable device can be produced which will acquire EEG signals and diagnose it automatically, without the help of an expert. It will finally, generate a detailed report, detailing the particulars of the EEG results of that particular sitting.

In this research work the data used has been taken out of the CHB-MIT Scalp EEG database. In the future, data from local medical facilities and local patients can be utilized to optimize the automated system for better performance. This may come out to be productive considering different atmospheric conditions as compared to those from where the data was actually taken.

The features used in this research work are from time, frequency and time-spectral domain. However, the spectral features used in this work are of slightly higher order. As the features of slightly higher order helped in improving the accuracy of the system, in future, more features of relatively higher order can be used for an improvement in the performance of the proposed system.

References

[1] Alexandros T. Tzallas, Epileptic Seizure Detection in EEGs Using Time-Frequency Analysis, IEEE Transaction on Information Technology in Biomedicine, Vol 13, No. 5, September 2009.

[2] Epilepsey Research and Treatment, Hindawi Publishing Corporation, http://www.hindawi.com/journals/ert/2012/205131

[3] Clinic for Epileptology, Department of Epileptology, University of Bonn, http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3

[4] Lowry, Richard. "Concepts & Applications of Inferential Statistics".Retrieved 24 March 2011.

[5] David F. Bauer (1972), Constructing confidence sets using rank statistics. .Journal of the American Statistical Association 67, 687-690.

[6]Brown S, Betts T, Crawford P et al. Epilepsy needs revisited: a revised epilepsy needs document for the UK. Seizure. 1998; 7(6):435-446.

[7] Kannathal, N., Choo, M.L., Acharya, U.R., Sadasivan, P.K.: Entropiesfor detection of epilepsy in EEG. Comput.Methods Programs Biomed.80, 187-194 (2005).

[8] Mitchell, Douglas W., "More on spreads and non-arithmetic means,"The Mathematical Gazette 88, March 2004, 142-144.

[9] Upton, Graham; Cook, Ian (1996). Understanding Statistics.Oxford University Press.p. 55.

[10] Kader, Gary (March 1999). "Means and MADS". Mathematics Teaching in the Middle School 4 (6): 398-403. Retrieved 20 February 2013.

[11] Susan Dean, Barbara Illowsky "Descriptive Statistics: Skewness and the Mean, Median, and Mode"

[12] .Ballanda, Kevin P.; MacGillivray, H. L. (1988)."Kurtosis: A Critical Review". The American Statistician (American Statistical Association) 42(2): 111-119

[13]http://www.biomedresearches.com/root/pages/researches/epilepsy/eeg_resources.html

[14] http://www.healthybrainforlife.com/images/brain/brain-lobes-color.gif/view

[15]http://en.wikipedia.org/wiki/Electroencephalography

[16]Ahmed Fazle Rabbiand Reza Fazel-Rezai, "A Fuzzy Logic System for Seizure Onset

Detection in Intracranial EEG", Computational Intelligence and Neuroscience, vol. 2012, Article ID 705140, 12 pages, 2012. doi:10.1155/2012/705140

[17] Sharanreddy and P.K. Kulkarni, "EEG Signal Classification for Epilepsy Seizure Detection usingImproved Approximate Entropy", International Journal of Public Health Science (IJPHS), Vol. 2, No. 1, March 2013, pp. 23~32ISSN: 2252-8806

[18] Yusuf U Khan, Omar Farooq and Priyanka Sharma, "Automatic Detection of Seizure Onset in Pediatric EEG", International Journal of Embedded Systems and Applications (IJESA) Vol.2, No.3, September 2012, p81-89. 16p

[19] Ali H Shoeb and John V Guttag, "Application of machine learning to epileptic seizure detection", Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pages 975-982

[20]Sachee, RoohiSille, Garima Sharma and N. Pradhan, "Implementation of Probabilistic Neural Network using Approximate Entropy to Detect Epileptic Seizures",

[21]Forrest Sheng Bao, Jue-Ming Gao, Jing Hu, Donald Y. C. Lie, Yuanlin Zhang, and K. J. Oommen, "Automated Epilepsy Diagnosis Using Interictal Scalp EEG", 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009

[22]A.S. MuthananthaMurugavel and S. Ramakrishnan, "Wavelet Domain Approximate Entropy-Based Epileptic Seizure Detection", ICIT 2011 The 5th International Conference on Information Technology

[23] Alison A. Dingle, Richard D. Jones, Grant J. Caroll and W. Richard Fright, "A Multistage System to Detect Epileptiform Activity in the EEG", IEEE Transactions on Biomedical Engineering, Vol. 40, No. 12, December 1993

[24] Alexandros T. Tzallas, Markos G. Tsipouras and Dimitrios I. Fotiadis, "Epileptic Seizure Detection in Electroencephalograms using Time-Frequency Analysis", Information Technology in Biomedicine, IEEE Transactions on, 13(5), 703-710.

[25] Lorena Orosco, Agustina Garcés Correa and Eric LaciarLeber, "Epileptic Seizures Detection Based on Empirical Mode Decomposition of EEG Signals", MANAGEMENT OF EPILEPSY–RESEARCH, RESULTS AND TREATMENT (2011): 1

[26] AbdulhamitSubasi and Ergun Ercelebi, "Classification of EEG signals using neural network and logistic regression", Computer Methods and Programs in Biomedicine (2005), Vol. 78, Issue 2, 87—99