

Analysis and Classification of EEG signals using Mixture of Features and Committee Neural Network

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

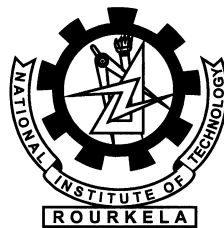
MASTER OF TECHNOLOGY

IN

ELECTRONICS AND INSTRUMENTATION ENGINEERING.

BY

Md Ashraf Jamal
Roll No: 210EC3287



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769008, ODISHA, INDIA
2012

Analysis and Classification of EEG Signals using Mixture of Features and Committee Neural Network

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

ELECTRONICS AND INSTRUMENTATION ENGINEERING.

BY

Md Ashraf Jamal
Roll No: 210EC3287

UNDER THE GUIDANCE OF

Dr. Samit Ari



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769008, ODISHA, INDIA
2012



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769 008, ODISHA, INDIA.

Certificate

This is to certify that the work in the thesis entitled “**Analysis and Classification of EEG signals using Mixture of Features and Committee Neural Network**” by Md **Ashraf Jamal** is a record of an original research work carried out by him during 2011 - 2012 under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of Master of Technology with the specialization of Electronics and Instrumentation Engineering in the department of Electronics and Communication Engineering, National Institute of Technology Rourkela. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Place: NIT Rourkela

Date: 1 June 2012

Dr. Samit Ari

Asst. Professor, ECE Department
NIT Rourkela, Odisha

Acknowledgments

Completion of this project and thesis would not have been possible without the help of many people, to whom I am very thankful. First of all, I would like to express my sincere gratitude to my supervisor, **Prof. Samit Ari**. His constant motivation, guidance and support helped me a great deal to achieve this feat.

I would like to thank Prof. S. K. Patra, Prof. K. K. Mahapatra, Prof. S. Meher, Prof. S. K. Behera, Prof. Poonam Singh and Prof. A. K. Sahoo for guiding and inspiring me in many ways. I am also thankful to other faculty and staff of Electronics and Communication department for their support.

I would like to mention the names of Manab, Dipak, Sharvan and Sumit and all other members of Computer Vision Lab for their constant support and co-operation throughout the course of the project. I would also like to thank all my friends within and outside the department for all their encouragement, motivation and the experiences that they shared with me.

I am deeply indebted to my parents who always had their belief in me and gave all their support for all the choices that I have made.

Finally, I humbly bow my head with utmost gratitude before the God Almighty who always showed me the path to go and without whom I could not have done any of these.

Md Ashraf Jamal

Contents

Certificate	ii
Acknowledgement	iii
Abstract	i
list of figures	iv
list of tables	v
1 Introduction	1
1.1 Introduction	2
1.1.1 Neural Activities	2
1.1.2 EEG signal generation	3
1.1.3 Recording of EEG signals	4
1.1.4 Frequency bands of EEG Signals	5
1.1.5 Advantages of EEG	6
1.1.6 Uses of EEG	6
1.1.7 Normal and abnormal EEG signals	6
1.2 Data Base	7
1.3 Objective	7
1.4 Thesis outline	8
2 Feature Extraction	10
2.1 Introduction	11
2.2 Autoregressive Process	11
2.2.1 Autoregressive Moving average process	11
2.2.2 Autoregressive process	12
2.2.3 Discrete Wavelet Transform	13
2.2.4 Wavelet Families	15

2.3	Feature Extraction	15
2.3.1	Normalization	16
2.3.2	Segment Detection	17
3	Classifier	26
3.1	Artificial neural Network	27
3.1.1	Introduction	27
3.1.2	Types of activation functions	28
3.1.3	Multilayer feedforward network	30
3.1.4	Learning process	31
3.1.5	Error-correction learning	32
3.1.6	Perceptron networks	33
3.1.7	Back-propagation networks	33
3.2	Committee Neural Network	34
3.3	Implementation of ANN and two-level CNN Classifier	35
3.3.1	Experiment using ANN classifier	37
3.3.2	Experiment using Two-level CNN classifier	37
4	Optimized Classifier	40
4.1	Introduction	41
4.2	Genetic Algorithm	41
4.2.1	Feature selection and Hidden node optimization	43
4.2.2	Experiment Using ANN and Committee Neural network with selected feature and optimized hidden neurons	46
5	Concluding Remarks	51
5.1	Conclusion	52
5.2	Future work	52
6	References	53

List of Figures

1.1	Structure of neuron [4]	3
1.2	Conventional 10-20 EEG electrode positions for the placement of 21 electrodes [4]	4
1.3	Raw EEG signal of class A, class D, and class E	8
2.1	Shape of (a) wave, (b) wavelet	14
2.2	Sub band decomposition of discrete wavelet transform	14
2.3	Wavelets functions (a) Haar, (b) Daubechies, (c) Coiflet1, (d) Symlet2, (e) Meyer, (f) Morlet, (g) Mexican Hat [21]	16
2.4	Flow diagram for classification of EEG signal	16
2.5	Raw EEG Normalization of class A, class D and class E	17
2.6	Raw EEG segment of class A, class D and class E	18
2.7	Autoregressive coefficient of class A, class D, and class E	19
2.8	power spectral density of class A	20
2.9	power spectral density of class D	20
2.10	power spectral density of class E	21
2.11	Plots of detailed and approximation coefficients of Class A, Class D, and Class E	22
2.12	Plots of 265 Wavelet coefficients of Class A, class D, and class E	23
2.13	Plots of 20 Wavelet coefficients of Class A, Class D, and Class E	24
2.14	Plots of mixture of features of Class A, Class D, and Class E	25
3.1	Threshold activation function	28
3.2	Piecewise-linear activation function	29
3.3	Sigmoid (logistic) activation function	29
3.4	Hyperbolic tangent activation function	30
3.5	Multilayer feedforward neural network	31

3.6	Error-correction learning diagram	32
3.7	A perceptron with many inputs and many outputs	33
3.8	Two-level Committee neural network	36
4.1	Flow diagram for classification of EEG signal	43
4.2	Representation of chromosome for feature Selection	44
4.3	Representation of chromosome for hidden node encoding	44
4.4	Representation of chromosome for feature selection and hidden node optimization	45
4.5	Flow chart of feature selection and hidden node optimization	47

List of Tables

2.1	Features extracted from 3 different recordings of 3 classes	21
3.1	Confusion matrix of ANN trained with AR features, DWT features, and mixture of features	37
3.2	Confusion matrix of individual member network of two-level CCN trained with AR features, DWT features, and mixture of features	38
3.3	Confusion matrix of CNN at level-1 and level-2 trained with AR features, DWT features, and mixture of features	38
3.4	Classification accuracy of ANN and two-level CNN classifier trained with Different kind of features	39
4.1	Confusion matrix of ANN with selected features and optimized hidden neurons	48
4.2	Confusion matrix of NN1, NN2, NN3, NW1, and NW2 with selected features and optimized hidden neurons	49
4.3	Confusion matrix of CNN-1 and CNN-2 with selected features and optimized hidden neurons	49
4.4	Selected features length and optimized hidden neurons	50
4.5	Accuracy matrix of different classifiers with full features and selected features	50
4.6	Comperission of proposed method with wavelet/mixture of experts and Com-bined neural network baed model	50

Abstract

Electroencephalography signal is the recording of electrical activity of brain, provides valuable information of the brain function and neurological disorder. this paper proposed committee neural network for classification of EEG signals. Committee neural network consists of different neural network that used multilayer perceptron back propagation algorithm. The number of input node and hidden node selection for artificial neural network remains an important issues, as over parametrized ANN gets trapped in local minima resulting non convergence of ANN structure during training. Redundant features and excessive hidden nodes of ANN increases modeling complexity without improving discrimination performance. Therefore optimum design of neural network which intern optimizes the committee neural network is required towards real time detection of EEG signals. The present work attempts to: (i) develop feature extraction algorithm which combines the score generated from autoregressive based feature and wavelet based feature for better classification of EEG signals, (ii) a two-level committee neural network is proposed based on the decision of several neural networks, (iii) select a set of input features that are effective for identification of EEG signal using genetic algorithm, (iv) make certain optimum selection of nodes in the hidden layer using genetic algorithm for each ANN structure of two-level CNN to get effective classification of EEG signal. It is observed that the performance of proposed technique is better than the earlier established techniques (combined neural network based model and wavelet/ mixture of experts network based approach) and the technique that uses artificial neural network with back propagation multilayer perceptron.

Chapter 1

Introduction

1.1 Introduction

Electroencephalography (EEG) signal is the recording of spontaneous electrical activity of the brain over a small period of time [1]. The term EEG refers that the brain activity emits the signal from head and being drawn. It is produced by bombardment of neurons within the brain. It is measured for a short duration of 20-40 minutes with the help of placing multiple electrodes over the scalp [2]. The researchers have found that these signals indicate the brain function and status of whole body. Therefore EEG signal provides valuable information of the brain function and neurobiological disorders as it provides a visual display of the recorded waveform and allows computer aided signal processing techniques to characterize them. This gives a prime motivation to apply the advanced digital signal processing techniques for analysis of EEG signals. EEG signal acquisition follows a noninvasive procedure [2] and proper signal processing and pattern recognition tools can emerge an important device for automated system to recognize electroencephalographic changes. Real time automated EEG signal analysis in clinical settings is of great assistance to clinicians in detecting neurobiological diseases and prevention of the possibility of missing (or misreading) information [3]. However, automated classification of EEG signals is a challenging problem as the morphological and temporal characteristics of EEG signals show significant variations for different patients and under different temporal and physical conditions. This ability motivates us to apply the advanced digital signal processing techniques on the EEG signals. The study of neuronal function and neurophysiological properties with EEG signals generation and their detection plays the vital role in detection, diagnosis of brain disorders and brain diseases.

1.1.1 Neural Activities

The Central Nervous System (CNS) consists of two types of cells, nerve cells and glia cells. All the nerve cell consists of axons, dendrites and cell bodies. The proteins developed in the cell body are transmitting information to other parts of the body. The long cylindrical shaped axon transmits the electrical impulse. Dendrites are connected either to the axons or dendrites of other inside cells and receive the electrical impulse from other nerves cells. The study of brain found that each nerve of human is approximately connected to 10000 other nerves [4]. The electrical activity is mainly due the current flow between the tip of dendrites and axons, dendrites and dendrites of cells. The level of these signals is in mV

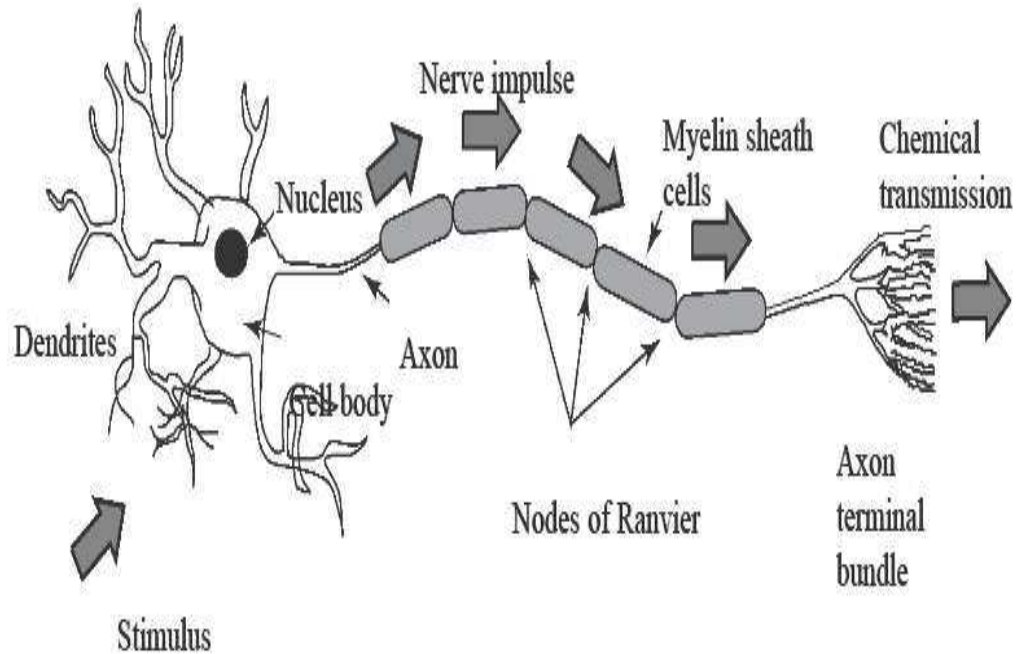


Figure 1.1: Structure of neuron [4]

range and its frequency is less than 100Hz [4].

1.1.2 EEG signal generation

The EEG signal is the current measured between the dendrites of nerve cells in the cerebrum region of the brain during their synaptic excitation. This current consists of electric field detected by electroencephalography (EEG) equipment and the magnetic field quantified by electromyogram (EMG) devices [4]. The human head consists of many parts some of which is scalp, skull, and brain. The noise is either to be produced internally in the brain or externally over the scalp due to the system used for recording. The degree of attenuation caused by skull is nearly hundred times larger than the soft tissues present in the head. As the level of signal amplitude is very low, the more number of neuron in excitation state can only produce the recordable potential by the scalp electrode system. Thus the amplifiers are

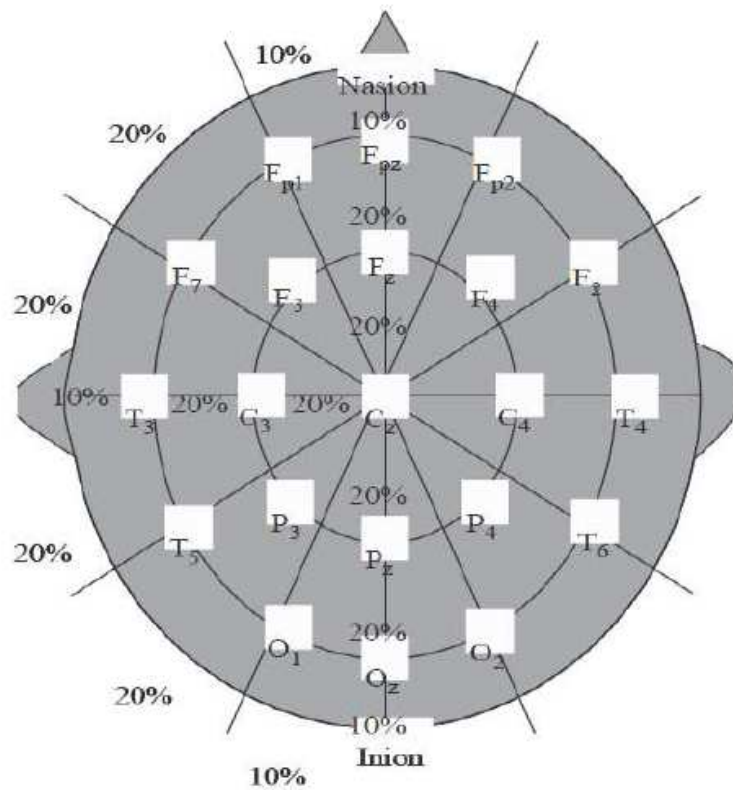


Figure 1.2: Conventional 10-20 EEG electrode positions for the placement of 21 electrodes [4]

used to increase the level of signals for later processing [5]. The brain structure is divided into three regions, cerebrum, cerebellum and brain stem. The cerebrum region defines the initiation of movement, conscious sensation and state of mind. The cerebellum region plays a role in voluntary actions like muscle movements. The brain stem region controls the respiration functioning, heart regulation, biorhythms and neural hormones [4], [6]. It is clear that the EEG signals generated from brain can determine the status of whole body and brain disorders [4], [6].

1.1.3 Recording of EEG signals

EEG is most often recorded from many electrodes in a arranged in a particular pattern or montage. A common standard for describing these position is the International 10/20 System. These methods are cheap and give a continuous record of brain activity with better

than millisecond resolution. No other tool can achieve this high temporal resolution and for this reasons many of the detailed discoveries of dynamic cognitive processes have been reported using EEG and ERP (Event Related Potentials) methods. The first electrical variation was noted by using a simple galvanometer. But in recent generation they are recorded by using EEG systems which consist of multiple electrodes, amplifiers single for each channel to amplify the attenuated signals and followed by filters to remove the system noise and registers [4], [6]. Later, EEG systems are equipped with computerized system to store the large data, for easy and correct analysis of EEG signals by using high sampling rate and more number of quantization levels with the help of advanced signal processing tools. The Analog to Digital Converter (ADC) circuits are used to convert the analog EEG signal to digital form. The bandwidth for EEG signal is 100 Hz. So, the minimum 200 samples/sec are required for sampling the EEG signal to satisfy the Nyquist criterion. Sometimes, the higher sampling rate of 2000 samples/sec is used for getting high resolution of EEG signals. The capacity of connected memory units depends upon the amount of data recorded. The different types of electrodes used for recording of high quality data.

1.1.4 Frequency bands of EEG Signals

Most of EEG waves range from 0.5-500 Hz, however following five frequency bands are clinically relevant: (i) delta, (ii) theta, (iii) alpha, (iv) beta and (v) gamma.

Delta waves: Delta waves frequency is up to 3 Hz. It is slowest wave having highest amplitude. It is dominant rhythm in infants up to one year and adults in deep sleep.

Theta waves: it is a slow wave having frequency range from 4 HZ to 7 Hz. It emerges with closing of the eyes and with relaxation. It is normally seen in young children or arousal in older children and adults.

Alpha waves: Alpha activity has frequency range from 7 HZ to 12 Hz. it most commonly seen in adults. Alpha activity occur rhythmically on both sides of the head but are often slightly higher in amplitude on the non dominant side, especially in right-handed individuals. Alpha wave appears with closing eyes (relaxation state) and disappears normally with opening eyes/stress. it is regarded as a normal waveform.

Beta waves: Beta activity is fast having small amplitude. It has frequency range from 14 HZ to 30 Hz. It is the dominant rhythm in patients who are alert or anxious or who have their eyes open. Beta waves usually seen on both sides in symmetrical distribution and is most evident frontally. It may be absent or reduced in areas of cortical damage. It

is generally regarded as a normal rhythm and observed in all age groups. These are mostly appeared in frontal and central portion of the brain. The amplitude of beta wave is less than $30 \mu\text{V}$.

Gamma waves: It is fastest waves of brain having frequency range falls from 30-45 Hz. It is also called as fast beta waves. These waves have very low amplitude and present rarely. But, the detection of these rhythms plays an important role in finding the neurological diseases. These waves are occurred in front central part of the brain. It suggests the event-related synchronization (ERS) of the brain [4].

1.1.5 Advantages of EEG

There are various advantage of EEG signals some of them can be stated as follows:

- Temporal resolution of EEG signal is high.
- EEG measure the electrical activity directly.
- EEG is a non-invasive procedure.
- EEG has ability to analyze brain activity.

Major disadvantage is EEG that it is difficult to find out the source of electrical activity from where it is coming out.

1.1.6 Uses of EEG

- Diagnose epilepsy and see what type of seizures are occurring. EEG is the most useful and important test in confirming a diagnosis of epilepsy.
- Check for problems with loss of consciousness or dementia.
- Help find out a person's chance of recovery after a change in consciousness.
- Find out if a person who is in a coma is brain-dead.
- Study sleep disorders, such as narcolepsy.
- Watch brain activity while a person is receiving general anesthesia during brain surgery.
- Help find out if a person has a physical problem (problems in the brain, spinal cord, or nervous system) or a mental health problem.

1.1.7 Normal and abnormal EEG signals

Normal EEG [7]: In adults who are awake, the EEG shows generally alpha waves and beta waves. The two sides of the brain show similar patterns of electrical activity. There

are no abnormal bursts of electrical activity and no slow brain waves on the EEG tracing. If flashing lights (photic stimulation) are used during the test, one area of the brain (the occipital region) have a brief response after each flash of light, but the brain waves are normal.

Abnormal EEG [7]: Different patterns of electrical activity are obtained from the two sides of brain. This represent a problem in one area or side of the brain is present. The EEG shows sudden bursts of electrical activity (spikes). These sudden changes may be caused by a brain tumor, injury, epilepsy. When a person has epilepsy, the location and exact pattern of the abnormal brain waves may help show what type of epilepsy or seizures the person has. In many people with epilepsy, the EEG may appear completely normal between seizures. The EEG records changes in the brain waves that may not be in just one area of the brain.

1.2 Data Base

The raw EEG signal is obtained from repository source as mentioned in [30] which consists of total 5 sets (classes) of data (SET A, SET B, SET C, SET D, and SET E) corresponding to five different pathological and normal cases. Three data sets are selected from 5 data sets in this work. These three types of data represent three classes of EEG signals (SET A contains recordings from healthy volunteers with open eyes, SET D contains recording of epilepsy patients in the epileptogenic zone during the seizure free interval, and SET E contains the recordings of epilepsy patients during epileptic seizures) as shown in Fig. 1.3. All recordings were measured using Standard Electrode placement scheme also called as International 10-20 system. Each data set contains the 100 single channel recordings. The length of each single channel recording was of 26.3 sec. The 128 channel amplifier had been used for each channel [37]. The data were sampled at a rate of 173.61 samples per second using the 12 bit ADC. So the total samples present in single channel recording are nearly equal to 4097 samples (173.61×23.6). The band pass filter was fixed at 0.53-40 Hz (12dB/octave) [38].

1.3 Objective

The main objective of our research is to analyze the acquired EEG signals using signal processing tools and classify them into different classes. The secondary goal is to improve

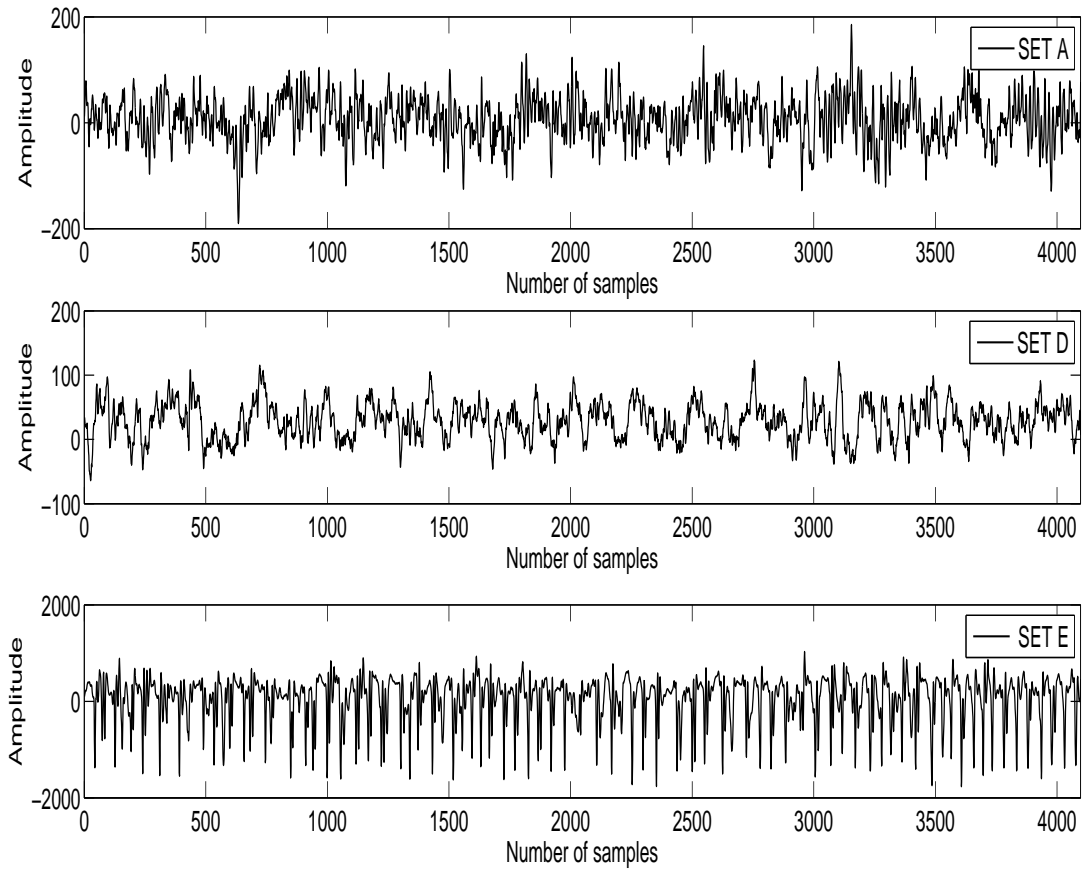


Figure 1.3: Raw EEG signal of class A, class D, and class E

the accuracy of classification. In order to achieve this: (i) first a feature extraction technique is proposed, which combines extracted coefficients of Daubechies-2 wavelet at fourth level decomposition and Burg's algorithm individually, (ii) a two-level committee neural network (CNN-2) is proposed for better classification of EEG signals. The final level decision i.e. the recognition accuracy of the classifier will be made by CNN-2,(iii) select a set of input features that are effective for identification of EEG signal using genetic algorithm,(iv) make certain optimum selection of nodes in the hidden layer using genetic algorithm for each ANN structure of two-level CNN to get effective classification of EEG signal.

1.4 Thesis outline

Chapter 1 of the thesis explains the basic theory behind the EEG signals, this chapter also discuss about the used data in the experiment.

Chapter 2 describe two feature extraction techniques (DWT coefficients and AR coefficients), this chapter also discuss about the proposed feature extraction technique (mixture of features).

Chapter 3 describes about Artificial neural network and the proposed two-level committee neural network.

In chapter 4 committee neural network is optimized using genetic algorithm by selecting the input feature and hidden nodes. This chapter also discuss the result of proposed classifier with the result of earlier established techniques.

Chapter 5 presents conclusion and future work. Chapter shows references.

Chapter 2

Feature Extraction

2.1 Introduction

Feature extraction is the collection of relevant information from the signal. Due to high transitions the signals are unable to distinct and insufficient to answer the status of individual. So, the DSP tools have been used for extracting the desired characteristics of different EEG signals and provided the knowledge of human state. A number of feature extraction methods have been present for study the EEG signal. The wavelet transform is used for extracting the feature vectors from EEG signal because of its ability to characterize time-frequency information which is important in this context [3], [12]- [13]. The model based approach give spectral analysis technique using Burg autoregressive (AR) method is also used to extract the features representing EEG signals [13]. The research work [13] demonstrated that the Burg's AR coefficients based features well represent the EEG signals. The benefits of the model based spectra are that it has good statistical stability for short segments of signal and has good spectral resolution which is less dependent on the length of the signal [13], [14], [15].

AR method is frequently used as model based method since AR parameters can be easily estimate by solving linear equations. [1], [8]- [9]. N. Hazarika *et. al.* reported a classification technique based on wavelet transform and artificial neural network [10]. *Übelli* [11] presented wavelet based feature set and a mixture of experts network for classification of three different pathological cases of 300 EEG recordings with accuracy 93.17 percent. In her another reported technique [12], a combined neural network with wavelet based feature set is used for classification of the same EEG data set and a performance of 94.8 percent is shown.

2.2 Autoregressive Process

2.2.1 Autoregressive Moving average process

An Autoregressive moving average process $x(n)$ can be generated by filtering the white noise $v(n)$, having zero mean and σ variance with a filter having p poles and q zeros

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^q b_k z^{-k}}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (2.1)$$

The power spectral density (PSD) of $x(n)$ is $p_x(z)$. In term of frequency variable w power spectral density is $p_x(e^{jw})$.

$$p_x(z) = \sigma^2 \frac{B_q(z)B_q^*(1/z^*)}{A_p(z)A_p^*(1/z^*)} \quad (2.2)$$

$$p_x(e^{jw}) = \sigma^2 \frac{|B_q(e^{jw})|^2}{|A_p(e^{jw})|^2} \quad (2.3)$$

A process having power spectral density in this form is called as an Autoregressive moving average process of order (p, q) and it is referred as ARMA (p, q) process. Power spectral density of this process having $2p$ poles and $2q$ zeros. Since $x(n)$ and $v(n)$ is related by the linear constant coefficient difference equation

$$x(n) + \sum_{l=1}^p a_l x(n-l) = \sum_{l=0}^p b_l v(n-l) \quad (2.4)$$

2.2.2 Autoregressive process

A special type of ARMA $(p, 0)$ process can be generated when $q=0$. In this case $x(n)$ is generated by filtering the white noise $v(n)$, with an all pole filter $H(z)$ of the form

$$H(z) = \frac{b(0)}{A_p(z)} = \frac{b(0)}{1 + \sum_{i=1}^p a_i z^{-i}} \quad (2.5)$$

An ARMA $(p, 0)$ process is called an autoregressive process of order p and it is referred as AR (p) process. $x(n)$ and $v(n)$ are related by the linear constant difference equation

$$x(n) + \sum_{i=1}^p a_i x(n-i) = b(0)v(n) \quad (2.6)$$

The power spectral density (PSD) of $x(n)$ is $p_x(z)$. In term of frequency variable w power spectral density is $p_x(e^{jw})$.

$$p_x(z) = \sigma^2 \frac{|b(0)|^2}{A_p(z)A_p^*(1/z^*)} \quad (2.7)$$

$$p_x(e^{jw}) = \sigma^2 \frac{|b(0)|^2}{|A_p(e^{jw})|^2} \quad (2.8)$$

To estimate the PSD, we required the estimation of variance (σ) and a_i coefficients. AR coefficient represent the power spectral density of the signal which carries important information for EEG classification [13]. Since the method characterizes the input data using an all-pole model, the correct choice of the model order p is important. In this context the value of model order can not be taken too large gives spurious peaks and too small result in smoothed estimate [13]. There are various methods available to estimates the reflection co-

efficient a_k , such as Autocorrelation (Yule-Walker), Covariance, and Burg's algorithm [16]. In this work burg method is used to estimates the reflection coefficient a_k is used to fit a p^{th} order autoregressive (AR) model to the input signal, x , by minimizing (least squares) the forward and backward prediction errors while constraining the AR parameters to satisfy the Levinson-Durbin recursion [16]. The model order of 10 is taken to get best performance [13].

2.2.3 Discrete Wavelet Transform

The signal can also be represented to another form by its transform by keeping the signal information to be same. The Short Time Fourier Transform (STFT) is used to study the non-stationary signals. To overcome the limit of its use to only non-stationary signals, the wavelet transform was developed to represent the signal into a function of time and frequency [17]. The STFT provides the identical resolution at all frequencies, but the use of wavelet transform provided a significance advantage to analyze the different frequencies of the signals with different resolutions by the method called multi-resolution technique [17]. So, this is perfect for analyzing the EEG signals [18]. The wavelets are the localized waves as their energy is concentrated in time which makes it best suitable for the analysis of transient signals [17]. The process of wavelet analysis is similar to that of STFT. In the first step of wavelet analysis, original signal is multiplied with the wavelet same as the signal is multiplied with window function in STFT case, and in the second step the transform is computed for each individual segment.

The wavelet transform provides multi-resolution description of a signal [11], [37]- [39]. It represents the problems of non-stationary signals in time frequency domain and hence is particularly suited for feature extraction of EEG signals [9]. Wavelet transform is computed using basis functions which are generated from the mother wavelet through scaling and translation operations [39]. It has a varying window size which is broad at low frequencies and narrow at high frequencies, thus providing optimal resolution at all frequencies [40]- [41]. The continuous wavelet transform is defined as

$$X_{WT} = \frac{1}{\sqrt{|s|}} \int x(t) \psi\left(\frac{t - \tau}{s}\right) dt \quad (2.9)$$

where $x(t)$ is the signal to be analyzed, $\Psi(t)$ is the mother wavelet or the basis function, τ is the translation parameter and s is the scale parameter [39]. The computation of continuous wavelet transform (CWT) consumes a lot of time and resources [42] and results in large amount of data, hence discrete wavelet transform (DWT), which is based on sub-band

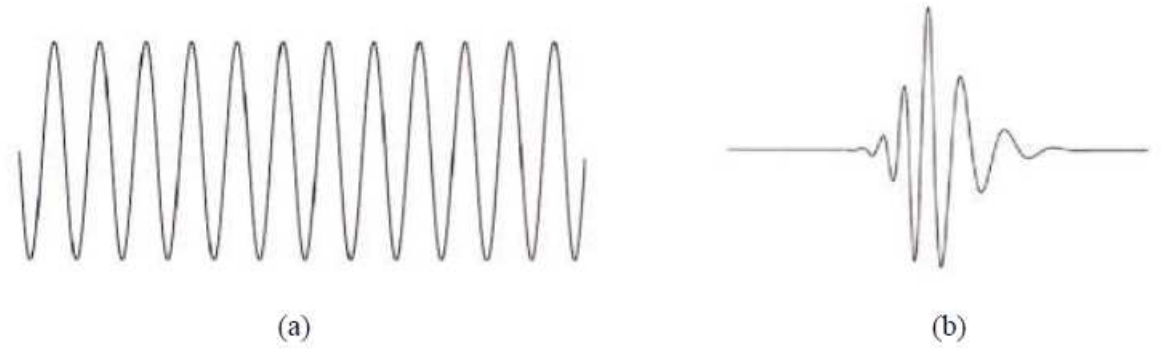


Figure 2.1: Shape of (a) wave, (b) wavelet

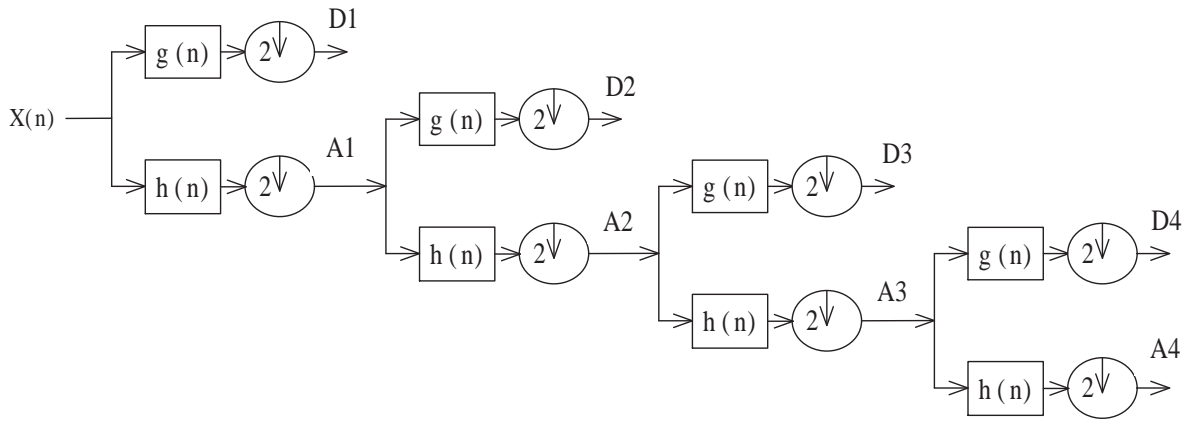


Figure 2.2: Sub band decomposition of discrete wavelet transform

coding is used as it gives a fast computation of wavelet transform. DWT can be described as a inner product between the given sequence and wavelet basis. DWT mathematically can be represented as [41]

$$W(j, k) = \sum_{n=0}^{N-1} x(n) \cdot \psi_{j,k}(n) \quad (2.10)$$

Where $W(j, k)$ is the DWT coefficients, $x(n)$ is the given sequence of length N , $\psi_{j,k}(n)$ is the wavelet basis

$$\psi_{j,k}(n) = \psi_{j,k}^* \left(\frac{n - s_o^j \cdot k}{s_o^j} \right) \quad (2.11)$$

$s_o^j, s_o^j k$ are discretized scale and translation parameter, $*$ denotes the complex conjugate.

In DWT the time-scale representation of the signal can be achieved using digital filtering

techniques, where basis function is represented by the set of high pass and low pass filter in the filter bank. The approach for the multi-resolution decomposition of a signal $x(n)$ is shown in Fig 2.2.

The DWT is computed by successive low pass and high pass filtering of the signal $x(n)$. Output of each filter is decimated by two. The high pass filter $g(n)$ is the discrete mother wavelet and the low pass filter $h(n)$ is its mirror version [1]. At each level, the down sampled outputs of the high pass filter produce the detail coefficients (high resolution component) D_j and that of low pass filter gives the approximation coefficients (low resolution component) A_j . The low resolution component are further decomposed and the procedure is continued up to four levels as shown [11], [37]- [38]. This filtering approach can be implemented by following equation [41]

$$D_{j+1}(i) = \sum_{l=0}^{L-1} g(l).A_j(2.i + l) \quad (2.12)$$

$$A_{j+1}(i) = \sum_{l=0}^{L-1} h(l).A_j(2.i + l) \quad (2.13)$$

where $j = 0, 1, 2, \dots, M$ are the decomposition levels. Initially A_0 is equals to original sequence $x(n)$. DWT has ability to extract the important information from EEG signal in time-frequency domain which helps the characterization and detection process [12], [43].

2.2.4 Wavelet Families

In wavelet transform basis function can be produced through translation and scaling the mother wavelet. As the mother wavelet produces all the basis functions which determine the characteristics of resultant wavelet transform, the selection of appropriate mother wavelet is necessary for finding of the wavelet transform effectively. The some of the commonly used wavelet functions are shown in Fig. 2.3.

2.3 Feature Extraction

Four different successive stage followed for classification of EEG signals are: normalization, segment detection, feature extraction, and classification. Fig. 4.1 shows the block diagram for decision making process..

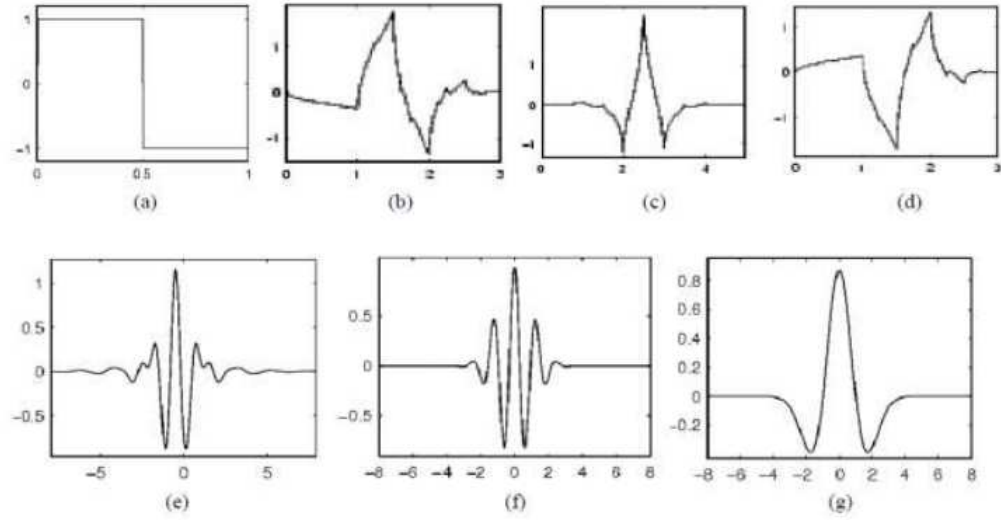


Figure 2.3: Wavelets functions (a) Haar, (b) Daubechies, (c) Coiflet1, (d) Symlet2, (e) Meyer, (f) Morlet, (g) Mexican Hat [21]

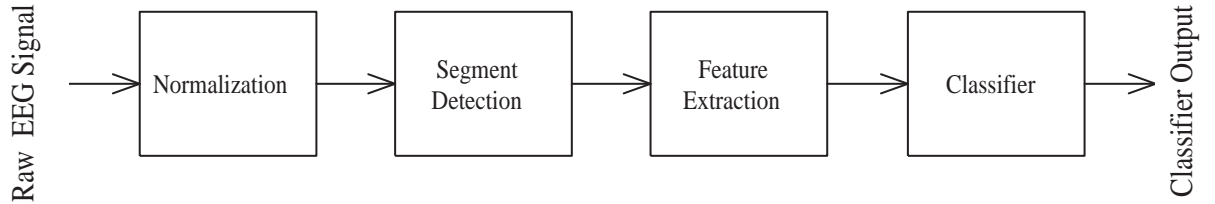


Figure 2.4: Flow diagram for classification of EEG signal

2.3.1 Normalization

Each recorded signals of all the 3 classes are normalized to -1 and +1 value. This is done by dividing all the samples with the maximum absolute value. The normalized plot of 3 class is shown in Fig. 2.5.

$$Normalize(k) = \frac{Signal(k)}{\|Signal(k)\|_{max}} \quad (2.14)$$

where $k = 1$ to 300.

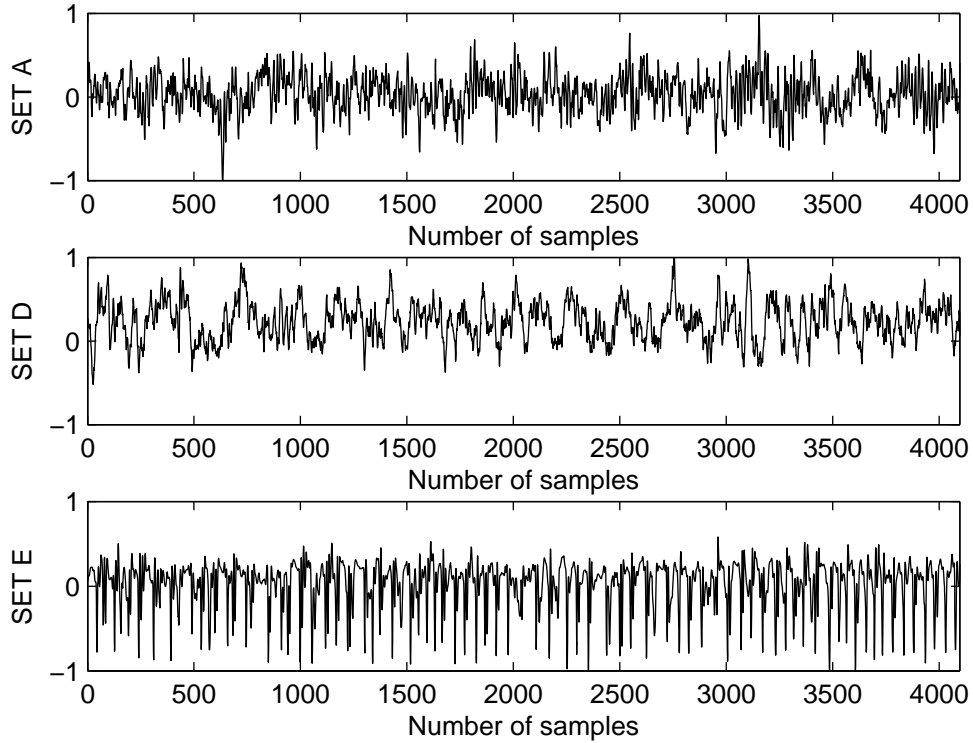


Figure 2.5: Raw EEG Normalization of class A, class D and class E

2.3.2 Segment Detection

A rectangular window of length 256 discrete samples is selected from each channel to form a single EEG segment. The total numbers of time series present in each class are 100 and each single channel time series contained 16 EEG signal segments. Therefore total 1600 EEG segments are produced from each class. Hence, total 4800 EEG segments are obtained from three classes. The 4800 EEG segments are divided into training and testing data sets. The 2400 EEG signal segments (800 vectors from each class) are used for testing and 2400 EEG signal segments (800 segments from each class) are used for training [11]. The plot of EEG segment from three classes (A, D, E) is shown in Fig. 2.6.

AR features: Autoregressive (AR) coefficients describe the important features of EEG signals, the correct choice of model order is important. Too low and too high order gives the poor estimation of power spectral density (PSD) [13]. In the present work, Burg's method is used for calculating the AR coefficients. The model order of 10^{th} is set to the input signals by minimizing the forward and backward prediction errors for the AR parameters to satisfy the Levinson-Durbin recursion. The vector of 11 AR coefficients is obtained for each

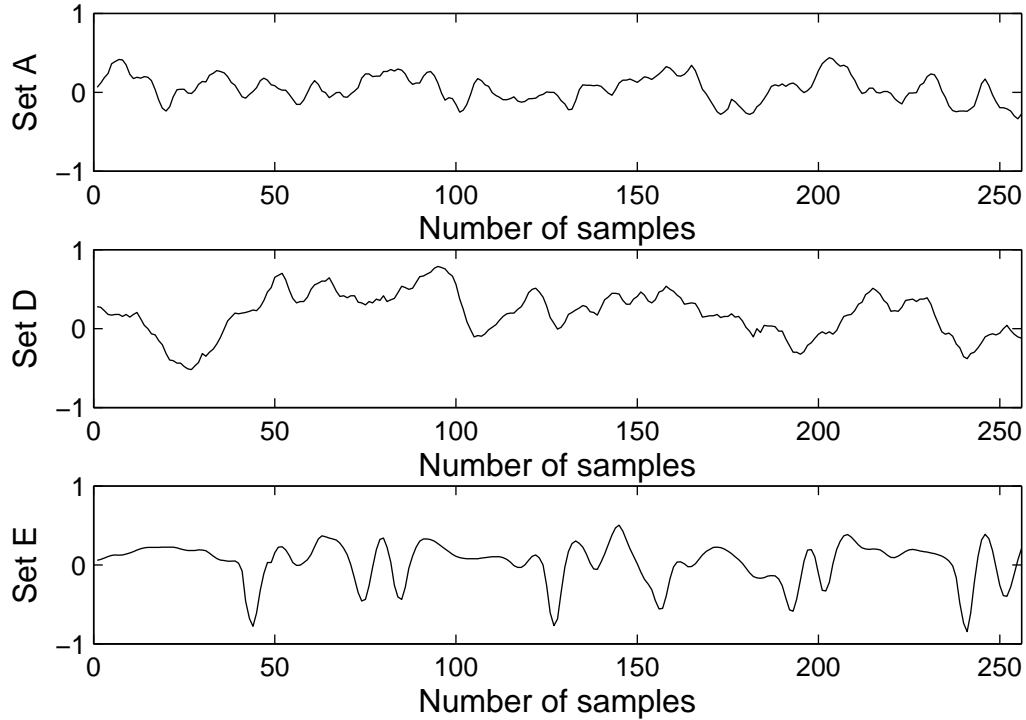


Figure 2.6: Raw EEG segment of class A, class D and class E

EEG segment is shown in the Fig. 2.7. The power spectral density (PSD) obtained by this method for the class A, class D, and class E are shown in Fig. 2.8, Fig. 2.9, and Fig. 2.10 respectively.

Wavelet Feature: For analysis of EEG signals using discrete wavelet transform, selection of the appropriate wavelet and number of decomposition level is important. The decomposition levels are picked depending upon main frequency components of the signal [1]. In the present work, daubechies wavelet of order 2 and 4 level of decomposition is used for computation of wavelet coefficients. The wavelet used here has smoothing characteristics which makes it suitable to detect changes in the EEG signal. Each of the EEG signals is decomposed into four detail wavelet coefficients (D1, D2, D3, and D4) and one approximation wavelet coefficients (A4). The representation D1, D2, D3, and D4 refers to the coefficients obtained at first, second, third and fourth level respectively. The plots of the coefficients at each level of decomposition for single EEG segment of three class is shown in . 2.11. The total 265 wavelet coefficients are computed for each EEG segment in which 247 are the detail coefficients ($129 + 66 + 34 + 18$) and 18 are the approximation coefficient [43]. The plot of 265 DWT coefficients is shown in the Fig. 2.12. The dimension of

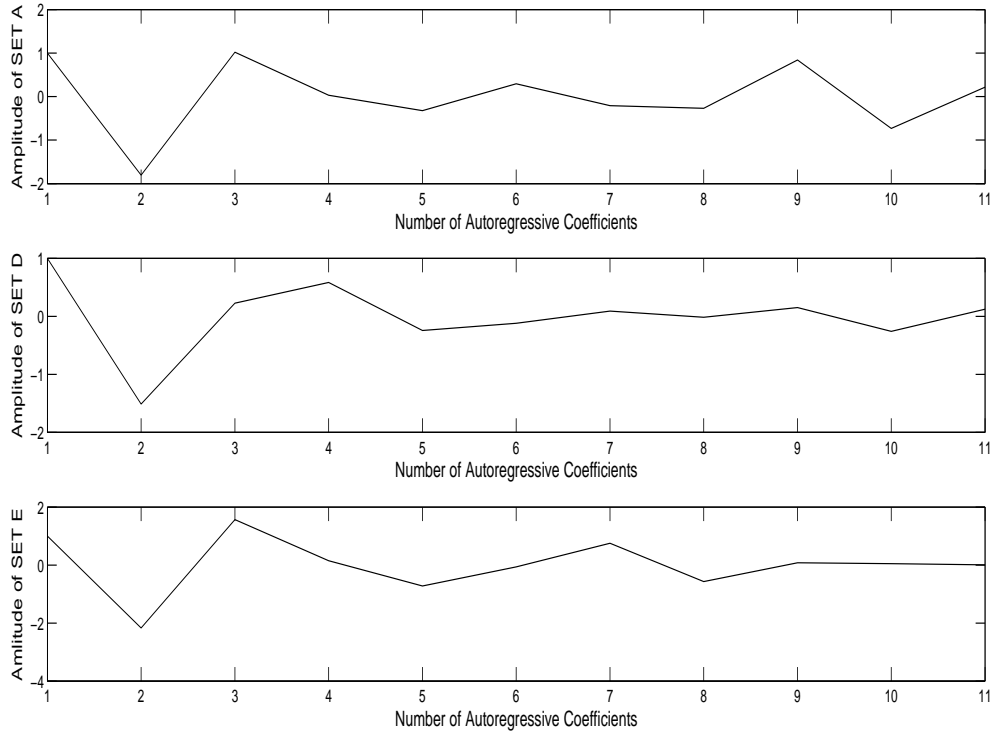


Figure 2.7: Autoregressive coefficient of class A, class D, and class E

the features obtained from DWT method is large. So different statistics methods are used to reduce the number of features are.

- (i) Maximum of wavelet coefficients in each sub band.
- (ii) Minimum of wavelet coefficients in each sub band.
- (iii) Mean of wavelet coefficients in each sub band.
- (iv) Standard Deviation of wavelet coefficients in each sub band [1].

Therefore, the 4 coefficients are obtained from each of the 5 sub bands resulting in a total of 20 coefficients feature vector as shown in Table 2.1. The plot of 20 dimension feature vectors for single EEG segment of Class A, Class D, and Class E is shown in Fig. 2.13

Mixture of features: This work proposed a mixture of features concept to represent EEG signals. Mixture of features are obtained by appending the 11 dimension AR coefficients feature vector with the 20 dimension DWT coefficients feature vector as a result 31 dimensional feature vector is produced. These 31 dimensional feature vectors are used for the proposed work is shown in Fig. 2.14.

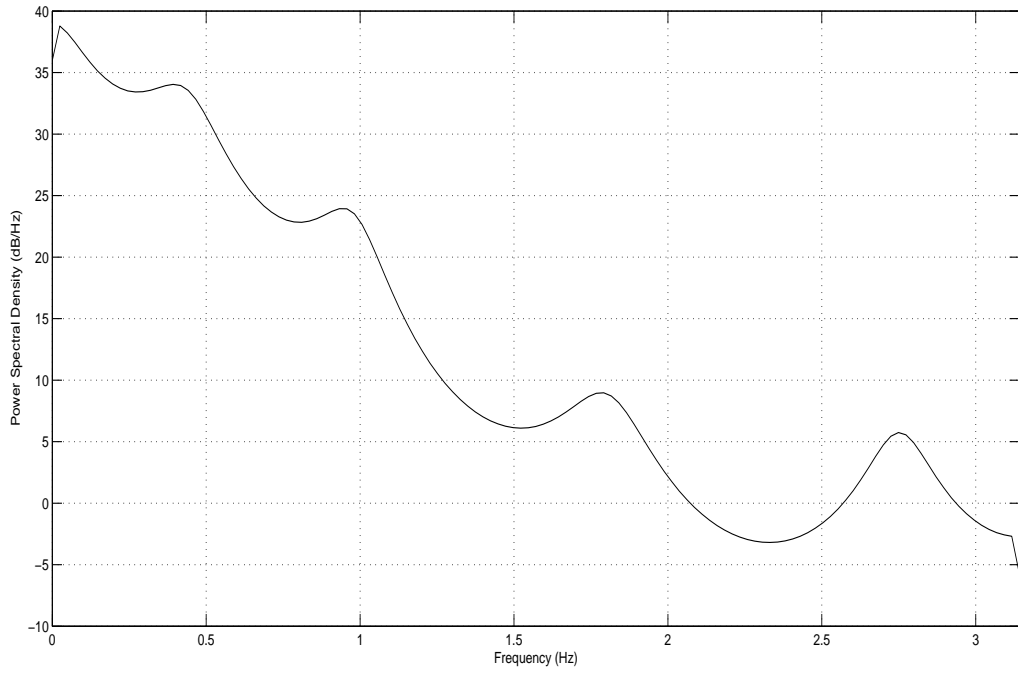


Figure 2.8: power spectral density of class A

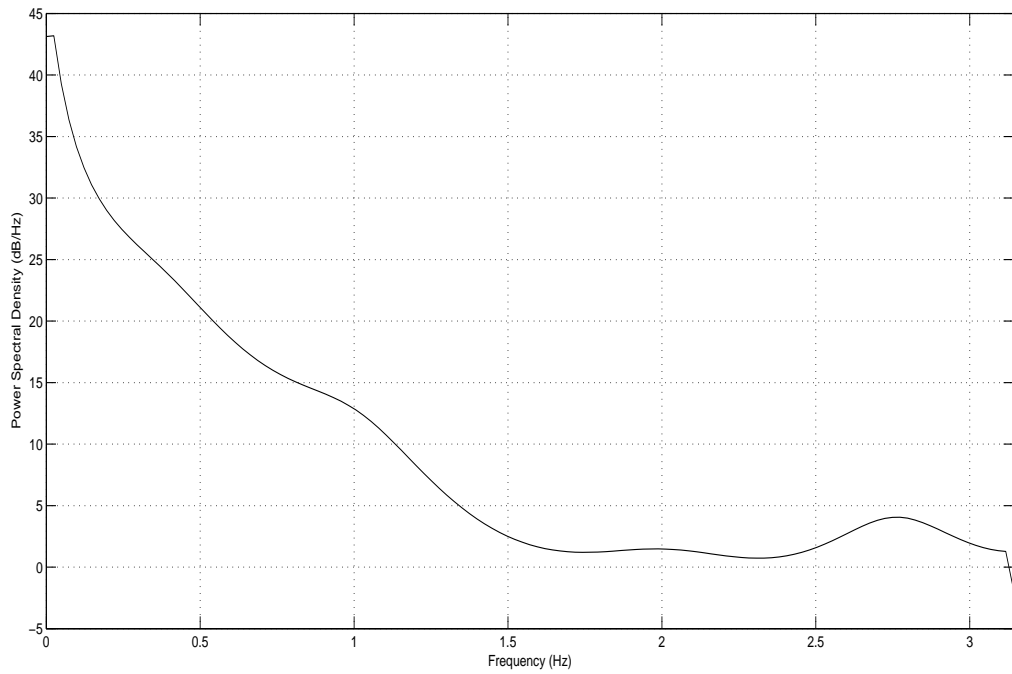


Figure 2.9: power spectral density of class D

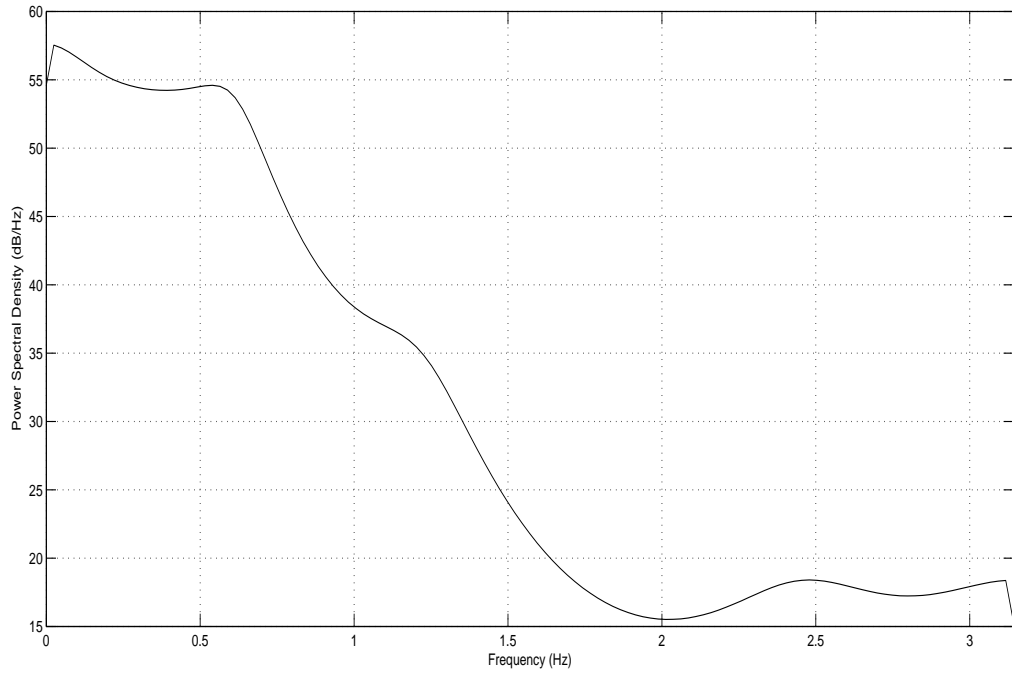


Figure 2.10: power spectral density of class E

Data set	Extracted features	sub-band				
		D1	D2	D3	D4	A4
Set A	Maximum	0.0633	0.1647	0.3987	0.6316	1.01408
	Minimum	-0.0632	-0.2214	-0.4861	-0.5545	-0.9078
	Mean	-0.0013	0.0009	0.0084	0.0114	0.1811
	Standard deviation	0.0261	0.0781	0.2167	0.3176	0.5076
Set D	Maximum	0.05363	0.194	0.3605	0.7174	2.6052
	Minimum	-0.0596	-0.1667	-0.2514	-0.7248	-1.429
	Mean	-0.0007	-0.00015	0.0134	-0.0214	0.7655
	Standard deviation	0.02374	0.07099	0.1581	0.3547	1.0272
Set E	Maximum	0.1462	0.365	0.8636	0.8045	0.9287
	Minimum	-0.1843	-0.6088	-0.8549	-0.6272	-1.0864
	Mean	0.00E+00	5.00E-05	0.0371	-0.0437	0.1594
	Standard deviation	0.04	0.172	0.4057	0.348	0.645

Table 2.1: Features extracted from 3 different recordings of 3 classes

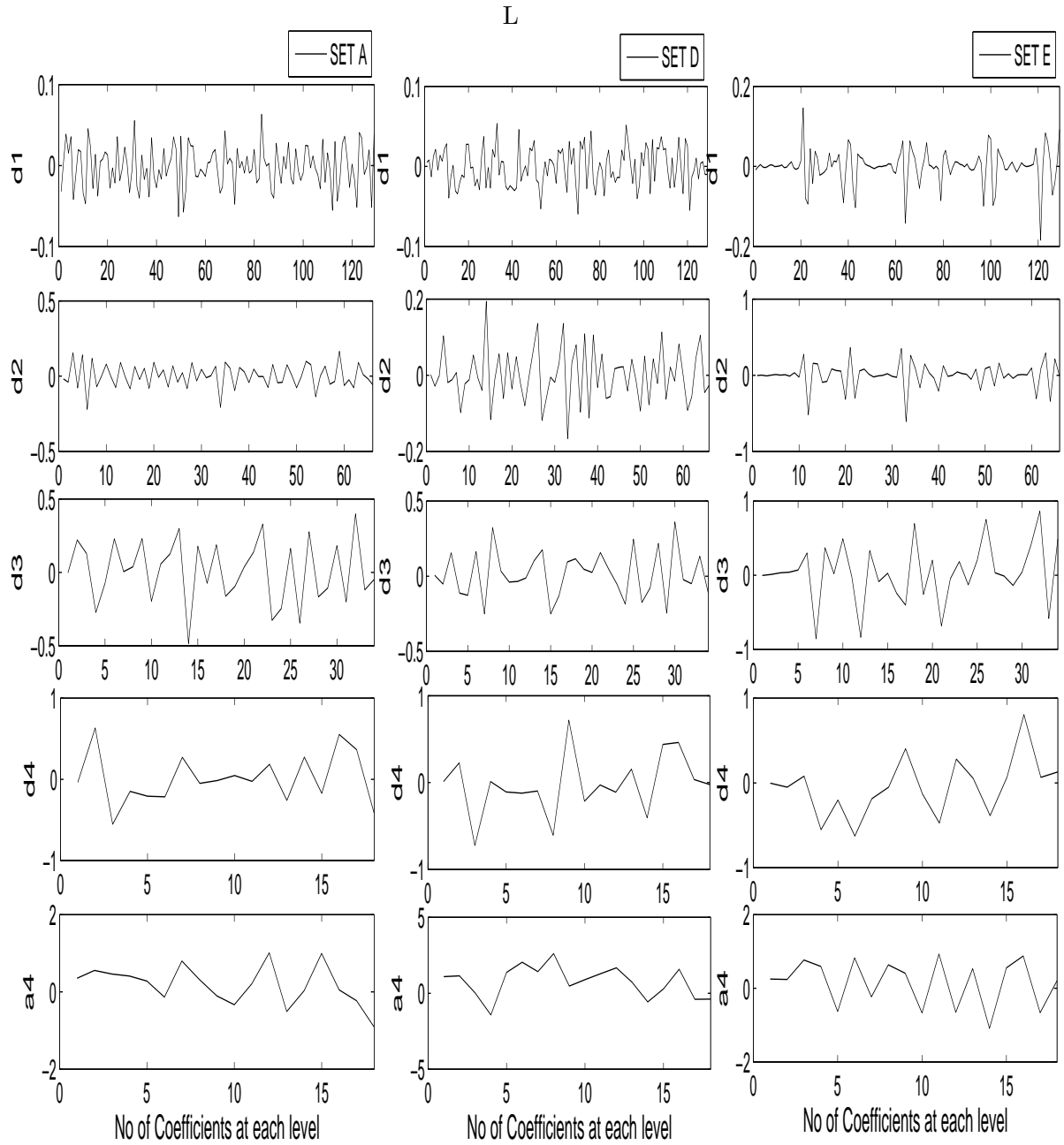


Figure 2.11: Plots of detailed and approximation coefficients of Class A, Class D, and Class E

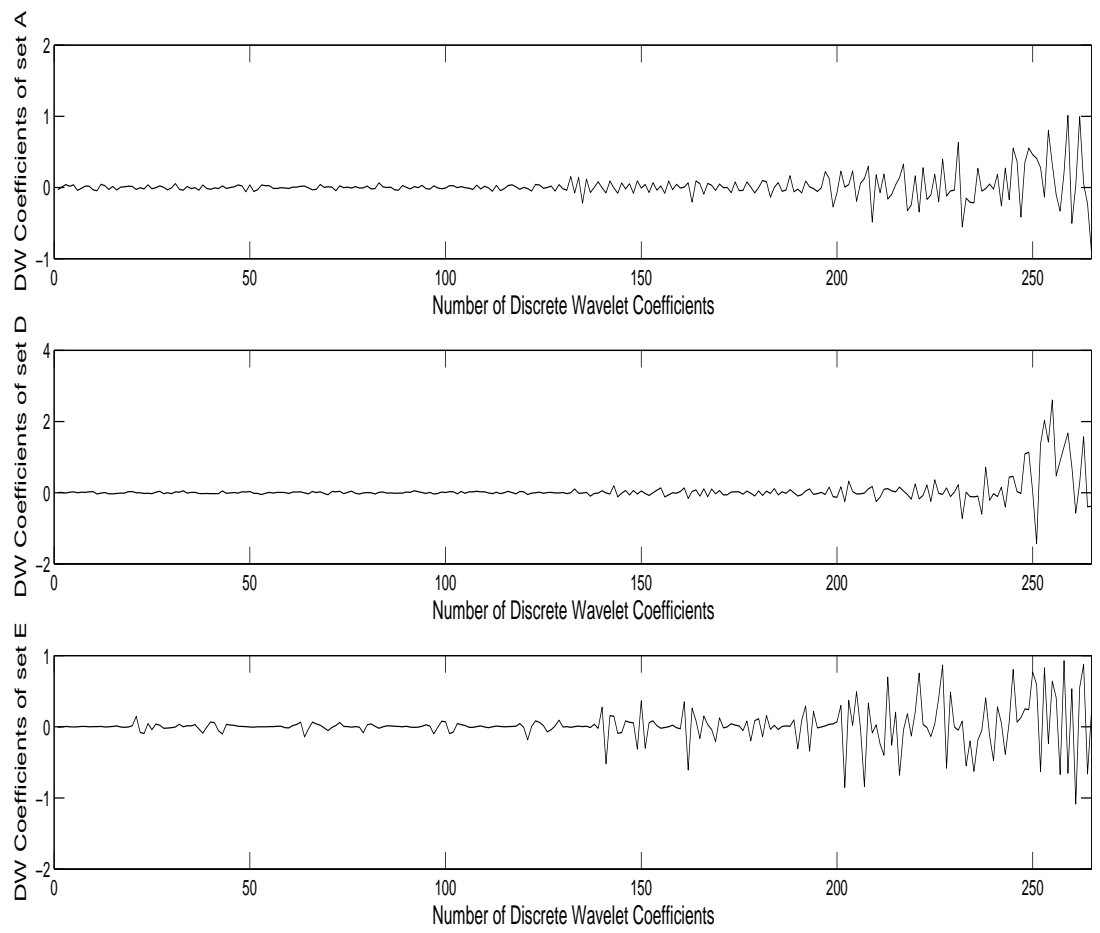


Figure 2.12: Plots of 265 Wavelet coefficients of Class A, class D, and class E

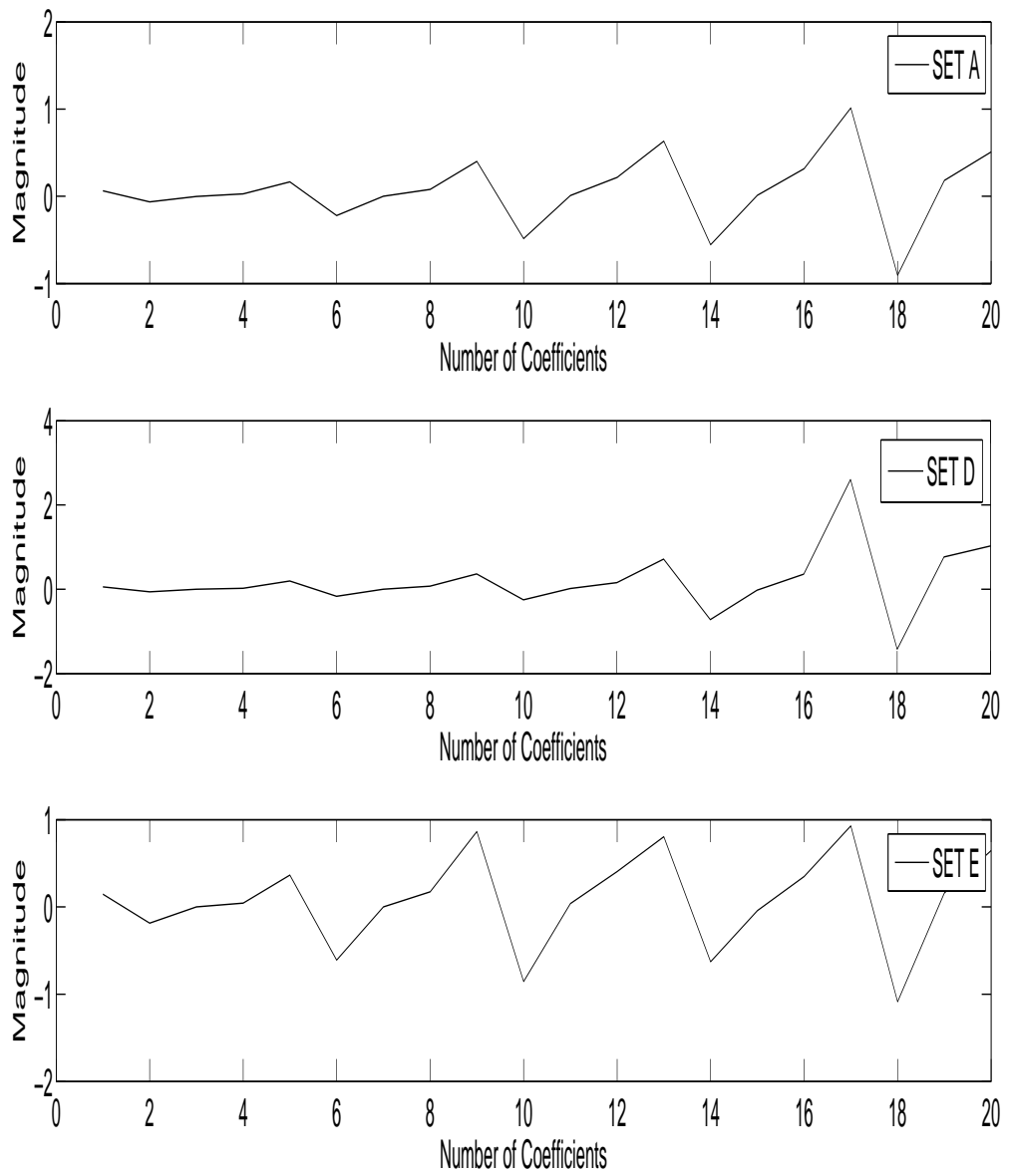


Figure 2.13: Plots of 20 Wavelet coefficients of Class A, Class D, and Class E

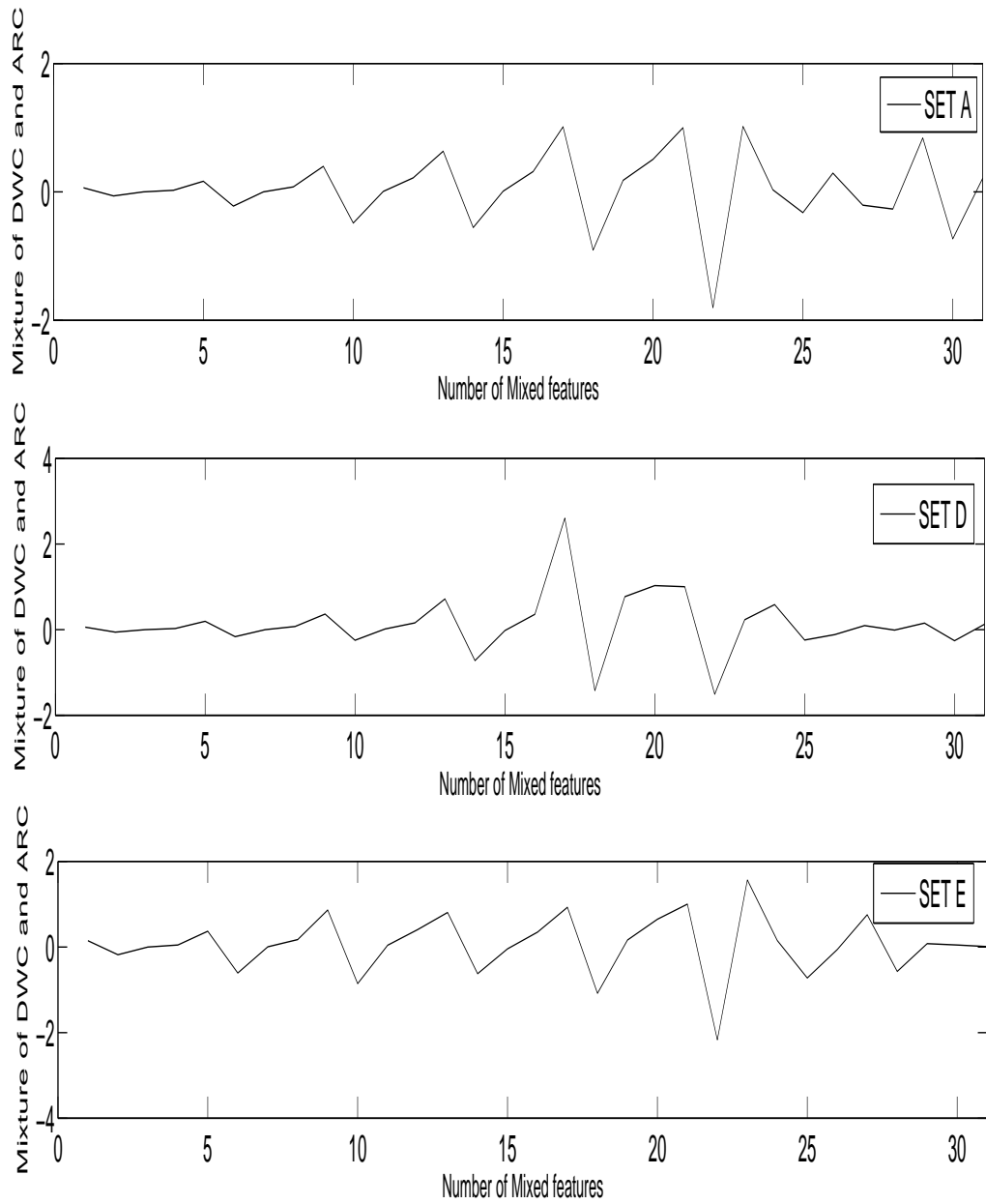


Figure 2.14: Plots of mixture of features of Class A, Class D, and Class E

Chapter 3

Classifier

3.1 Artificial neural Network

3.1.1 Introduction

Artificial neural networks is a mathematical model inspired by the functional aspect of biological structure of neurons commonly known as neural networks.

Artificial Neural Network (ANN) is normally used as a classifier [16]- [21] to discriminate the EEG signals. The popularity of ANN as a classifier is because (i) ANN can be used to generate likelihood-like scores that are discriminative in the state level; (ii) ANN can be easily implemented in hardware platform for its simple structure; (iii) ANN has the ability to approximate functions and automatic similarity based generalization property; (iv) complex class distributed features can be easily mapped by ANN [22]. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. ANN can be modeled on a human brain [35]. The basic processing unit of brain is neuron which works identically in ANN. The neural network is formed by a set of neurons interconnected with each other through the synaptic weights. It is used to acquire knowledge in the learning phase. The number of neurons and synaptic weights can be changed according to desired design perspective [35]. A neuron is an information-processing unit that is fundamental to the operation of a neural network. We define the three basic elements of the neuron model:

1. A signal x_j at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj} . The weight w_{kj} is positive if the associated synapse is excitatory, it is negative if the synapse is inhibitory.
2. An adder of summing the input signals, weighted by the respective synapses of the neuron.
3. An activation function for limiting the amplitude of the output of a neuron. Typically, the normalized amplitude range of the output of a neuron is written as the $[0, 1]$ or alternatively $[-1, 1]$.

The model of a neuron also includes an externally applied bias (threshold) $w_{k0} = b_k$ that has the effect of lowering or increasing the net input of the activation function. In mathematical terms, we can describe a neuron k by writing the following pair of equations:

$$v_k = \sum_{j=0}^p w_{kj} x_j \quad (3.1)$$

$$y_k = \phi(v_k) \tag{3.2}$$

or in a matrix form

$$v_k = [w_{k0} \ w_{k1} \ \dots \ w_{kp}] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \end{bmatrix} = w_k^T x \tag{3.3}$$

3.1.2 Types of activation functions

Threshold activation function (McCullochPitts model)

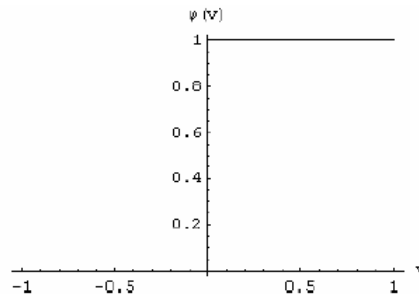


Figure 3.1: Threshold activation function

Output of a neuron takes the value 1 if the total internal activity level of that neuron is nonnegative and 0 otherwise. This statement describes the all-or-none property of the McCullochPitts model.

$$\varphi(v) = \begin{cases} 1, & v \geq 0 \\ -1, & v < 0 \end{cases} \tag{3.4}$$

Piecewise-linear activation function

$$\varphi(v) = \begin{cases} 1, & v \geq 1/2 \\ v + 1/2, & -1/2 < v < 1/2 \\ 0, & v \leq -1/2 \end{cases} \tag{3.5}$$

The amplification factor inside the linear region is assumed to be unity. The following two situations may be viewed as special forms of the piecewise linear function:

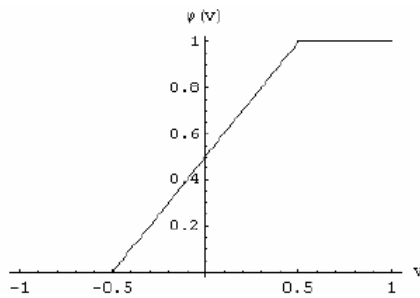


Figure 3.2: Piecewise-linear activation function

1. A *linear combiner* arises if the linear region of operation is maintained without running into saturation.
2. The piecewise-linear function reduces to a threshold function if the amplification factor of the linear region is made infinitely large.

Sigmoid (logistic) activation function

$$\varphi(v) = \frac{1}{1 + \exp(-v)} \quad (3.6)$$

In the construction of artificial neural networks most common activation function used is

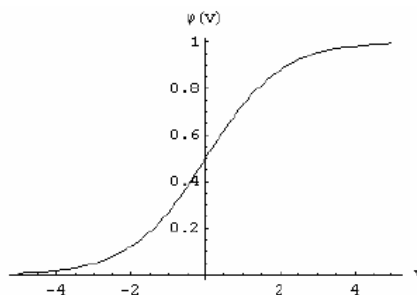


Figure 3.3: Sigmoid (logistic) activation function

sigmoid function. A sigmoid function has a range 0 and 1. Sigmoid function is differen-

tiable, which makes it easy to calculate.

$$\frac{\delta\varphi(v)}{\delta v} = \varphi(v)(1 - \varphi(v)) \quad (3.7)$$

Hyperbolic tangent function

$$\varphi(v) = \tanh\left(\frac{v}{2}\right) = \frac{1 - \exp(-v)}{1 + \exp(-v)} \quad (3.8)$$

The hyperbolic tangent function can be expressed in terms of the logistic function: $(2 \times$

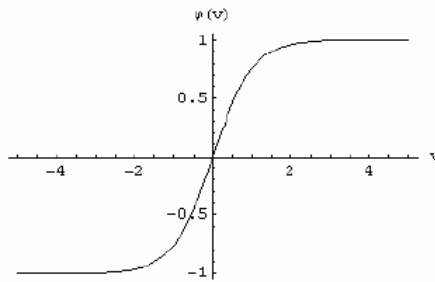


Figure 3.4: Hyperbolic tangent activation function

logistic function $\frac{1}{2}$). Its derivative is also easy to calculate:

$$\frac{\delta\varphi(v)}{\delta v} = \frac{1}{2}(1 + \varphi(v))(1 - \varphi(v)) \quad (3.9)$$

3.1.3 Multilayer feedforward network

The input-layer of the network supply the input signals (input vector) to the neurons in the second layer (i.e. the *first hidden layer*). The output of the second layer acts as an inputs to the third layer, and so on for the rest of the network. Generally, each layer's neurons have as their inputs output signals of the preceding layer only. The output signals of the neurons in the output layer of the network constitutes the overall response of the network to the activation pattern supplied by the source nodes in the input layer. A feed-forward network with p source nodes, h_1 neurons in the first hidden layer, h_2 neurons in the second layer, and q neurons in the output layer is referred to as a p - h_1 - h_2 - q network.

A neural network is said to be *fully* connected if every node in each layer of the network

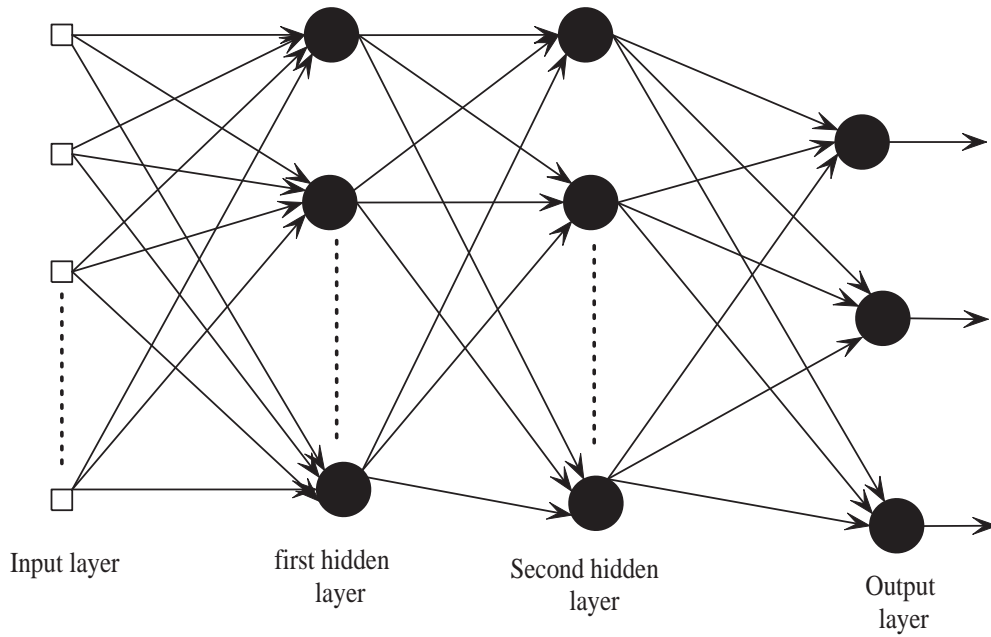


Figure 3.5: Multilayer feedforward neural network

is connected to every other node in the adjacent forward layer. If, however, some of the communication links (synaptic connections) are missing from the network, we say that the network is *partially connected*.

3.1.4 Learning process

Among the many interesting properties of a neural network is the ability of the network to learn from its environment and to improve its performance through learning. A neural network learns about its environment through an iterative process of adjustments applied to its synaptic weights and thresholds. The type of learning is determined by the manner in which the parameter changes take place.

Let $w_{kj}(n)$ is the value of the synaptic weight w_{kj} at time n . At time n an adjustment $\Delta w_{kj}(n)$ is applied to the synaptic weight $w_{kj}(n)$, yielding the updated value

$$w_{kj}(n + 1) = w_{kj}(n) + \Delta w_{kj}(n) \tag{3.10}$$

A prescribed set of well-defined rules for the solution of a learning problem is called a *learning algorithm*. Learning algorithms differ from each other in the way in which the adjustment Δw_{kj} to the synaptic weight w_{kj} is carried out.

3.1.5 Error-correction learning

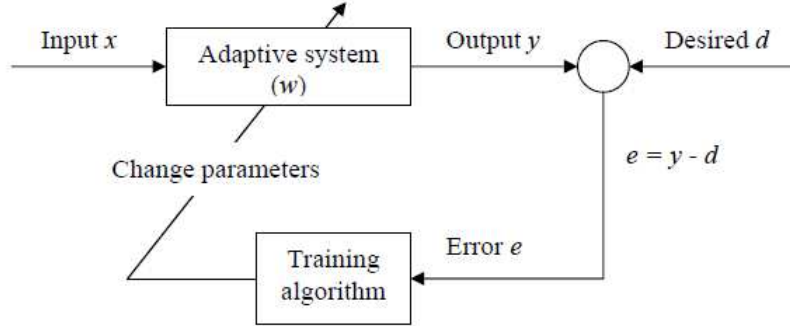


Figure 3.6: Error-correction learning diagram

Let $d_k(n)$ denote some desired response or target response for neuron k at time n . Let the corresponding value of the actual response (output) of this neuron be denoted by $y_k(n)$. The response $y_k(n)$ is produced by a stimulus applied to the input of the network in which neuron k is embedded. The input vector and desired response $d_k(n)$ constitute a particular sample presented to the network at time n .

Actual response $y_k(n)$ of neuron k is different from the desired response $d_k(n)$. Hence, we may define an error signal

$$e_k(n) = y_k(n) - d_k(n) \quad (3.11)$$

The ultimate purpose of error-correction learning is to minimize a cost function based on the error signal $e_k(n)$.

A criterion commonly used for the cost function is the instantaneous value of the mean-square-error criterion

$$J(n) = \frac{1}{2} \sum_k e_k^2(n) \quad (3.12)$$

The network is then optimized by minimizing $J(n)$ with respect to the synaptic weights of the network. Thus, according to the *error-correction learning rule* (or *delta rule*), the synaptic weight adjustment is given by

$$\Delta w_{kj} = \eta e_k(n) x_j(n) \quad (3.13)$$

3.1.6 Perceptron networks

Perceptron is the simplest form of a neural network used for the classification of a linearly separable patterns. It consists of a single neuron with adjustable synaptic weights and bias. It has been proved that if the perceptron are used to train the linearly separable classes, then the perceptron algorithm converges and positions the decision surface in the form of a hyperplane between the classes. The proof of convergence of the algorithm is known as the *perceptron convergence theorem*.

The single-layer perceptron shown has a single neuron. Such a perceptron is limited to performing pattern classification with only two classes. Several perceptrons can be combined to compute more complex functions.

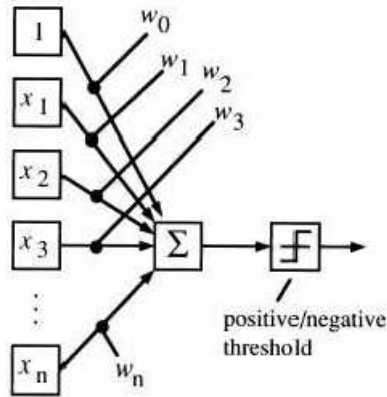


Figure 3.7: A perceptron with many inputs and many outputs

3.1.7 Back-propagation networks

Multilayer perceptrons have been applied successfully to solve some difficult diverse problems by training them in a supervised manner with a highly popular algorithm known as the *error back-propagation algorithm*. This algorithm is based on the *error-correction learning rule*.

The error back-propagation involves two steps- forward and backward passes through the different layers of the network. In the forward pass, The feature pattern is applied to the input in the forward direction and this input propagates by performing computations at each layer. Finally, a set of outputs is produced as the actual response of the network. This

principle of this algorithm is based on delta rule and gradient descent [19]. In the delta rule, the weights change of a neuron is proportional to the learning rate parameter and gradient function. The synaptic weights remain same in the forward pass. In the output layer the out value corresponding to the particular input is acquired and the error is computed between the resulting output and target value. This error is passed in backward direction with the computation of local gradient at each layer and changes the weights to get the minimum error for desired response. This forward and backward process run repeatedly till the goal is achieved [19].

3.2 Committee Neural Network

Committee neural network is a classifier that reaps the benefits of its individual neural network members. It produces output [19], [20] by taking major decision of each individual neural network. In paper [21] single level committee neural network was used for the verification of speech signal. In this work, two- level committee neural networks has been proposed for classification of EEG signals. The block diagram of proposed Two-level committee neural network is shown in Fig. 3.8. First level of committee (CNN-1) is formed by using three different neural networks (NN1, NN2, and NN3) that used MLP-BP algorithm. The second level of committee (CNN-2) is formed by the output of the first level (CNN-1) and two other different neural networks (NW1 and NW2) that also used MLP-BP algorithm. Therefore total five member neural networks are used for two-level CNN. The training set for all the networks are different. Out of the different networks employed for the initial training stage the best performing three networks were selected to form the committee. In paper [21] single level committee neural network was used for the verification purpose. Here committee has been used for the classification purpose in which majority decision of the indivisual member network form the final decision of committee. Final level decision will be taken by the CNN-2, i.e. final output will be decided by the score generated NW1, NW2 and decision taken by CNN-1.

Each neural network is trained independently with different amount of training data and tested with with the same amount of testing data. Output of CNN-1 (level-1 Decision) is the majority decision formed by output of networks NN1, NN2, and NN3. Final output, i.e, output of CNN-2 (level-2 Decision) is the majority decision taken by the output of networks NW1, NW2 and CNN-1. The block diagram of proposed Two-level committee neural

network is shown in Fig. 3.8.

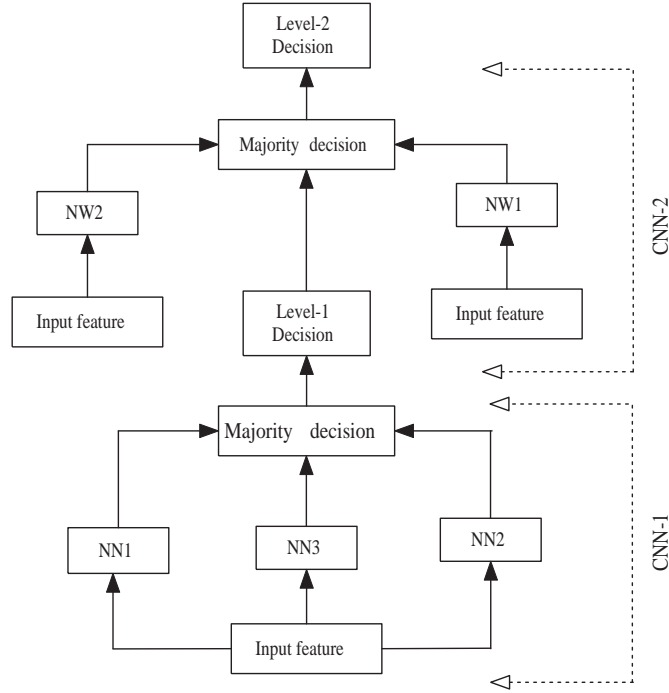


Figure 3.8: Two-level Committee neural network

3.3 Implementation of ANN and two-level CNN Classifier

The performance of proposed method is evaluated on three-class EEG signal data set which are publically available [30]. Total 4800 segments (1600 segments from each set) are obtained from three classes. Out of 4800 segments, 2400 segments (800 segment of set A, 800 segment of set D, and 800 segment of set E) are used as a training data set and rest 2400 segments (800 segments from each set) are used for testing data set.

Each member neural network of CNN is trained with different set of training data set. NW1 is trained using 1200 (first 400 segments from each training set) segments. NW2 is trained using 1200 (last 400 segments from each training set) segments. NN1 is trained using 768 (first 256 segments from each training set) segments. NN2 is trained using 768 (next 256 segments from each training set) segments. NN3 is trained using 864 (last 288 segments from each class) segments.

In order to compare the performance of proposed CNN for the same classification problem, we also implemented the ANN that used MLP-BP algorithm having 2400 training

data set (800 segments from each set). Each single hidden layer feedforward neural network (ANN, NW1, NW2, NN1, NN2, and NN3) is being tested using 2400 segments (800 segments from each set). Each neural network (ANN, NW1, NW2, NN1, NN2, and NN3) is being trained independently with the input feature vectors, where initial weight of the neural network is taken randomly. The number of hidden neurons, as well as the appropriate values of learning rate and momentum are estimated using a trial-and-error process until no further improvement in classification could be obtained. Performance of the classifier is evaluated based on the following statistical parameters [11]:

Specificity: Number of correct classified healthy segments divided by number of total healthy segments.

Sensitivity (seizure free epileptogenic zone segments): Number of correct classified seizure-free epileptogenic zone segments divided number of total seizure-free epileptogenic zone segments.

Sensitivity (epileptic seizure segments): Number of correct classified epileptic seizure segments divided by total number of epileptic seizure segments.

Accuracy: number of correct classified segments divided by number of total segments.

3.3.1 Experiment using ANN classifier

Classifier	Desire Result	Output								
		AR			DWT			Mixture of feature		
		SET A	SET D	SET E	SET A	SET D	SET E	SET A	SET D	SET E
ANN	SET A	795	21	6	725	63	54	789	28	8
	SET D	5	732	24	20	688	59	11	768	17
	SET E	0	47	770	55	49	687	0	34	775

Table 3.1: Confusion matrix of ANN trained with AR features, DWT features, and mixture of features

Neural network ANN is trained with the input Feature obtained by DWT, AR and mixture of features. Classification accuracy obtained with these three methods is shown in Table 3.1. From Table 3.4 it is verified that the mixture of feature provides the more accurate result compare to single alone feature. Therefore mixture of feature well represent and characterize the EEG signal more effectively than using any of them alone.

3.3.2 Experiment using Two-level CNN classifier

Classifier	Desire Result	Output								
		AR			DWT			Mixture of feature		
		SET A	SET D	SET E	SET A	SET D	SET E	SET A	SET D	SET E
NN1	SET A	784	50	21	678	107	86	667	34	26
	SET D	14	701	102	36	607	78	24	730	64
	SET E	2	49	677	86	54	668	9	36	710
NN2	SET A	744	25	9	653	78	48	746	26	13
	SET D	53	723	44	43	666	138	48	746	33
	SET E	3	52	747	105	56	614	66	28	754
NN3	SET A	725	29	12	680	52	76	781	41	21
	SET D	21	654	6	45	624	60	10	678	4
	SET E	4	117	782	75	124	658	9	81	775
NW1	SET A	758	41	18	701	105	70	772	32	0
	SET D	36	713	56	17	620	80	22	748	39
	SET E	6	46	726	82	75	650	66	20	761
NW2	SET A	757	16	9	637	4	70	789	64	27
	SET D	37	684	28	113	678	119	11	693	23
	SET E	6	100	768	50	81	611	0	43	750

Table 3.2: Confusion matrix of individual member network of two-level CCN trained with AR features, DWT features, and mixture of features

To form the proposed two-level committee neural networks five individual member neural network are trained by corresponding training data obtained from DWT, AR and appended features (mixture of features) having different hidden neurons. Table 3.2 represents the confusion matrix of member neural networks of tow-level CNN. Out put of these networks (NW1, NW2, NN1, NN2, and NN3) are applied to form committee neural network as shown in Fig. 3.8. Performance of proposed committee network at level-2 (CNN-2) are shown in the Table 3.4.

Classifier	Desire Result	Output								
		AR			DWT			Mixture of feature		
		SET A	SET D	SET E	SET A	SET D	SET E	SET A	SET D	SET E
CNN-1	SET A	787	23	12	707	61	41	788	18	16
	SET D	8	711	16	21	658	58	9	745	12
	SET E	5	66	772	72	81	701	3	37	772
CNN-2	SET A	756	16	9	711	51	40	794	20	7
	SET D	36	681	9	24	676	56	6	758	10
	SET E	8	103	782	65	73	704	0	22	783

Table 3.3: Confusion matrix of CNN at level-1 and level-2 trained with AR features, DWT features, and mixture of features

Classifier	Accuracy (in percentage)		
	Feature1 (DWT)	Fature2 (AR)	Mixture of feature
ANN	85.5	94.91	95.91
Proposed CNN	86.90	94.70	97.29

Table 3.4: Classification accuracy of ANN and two-level CNN classifier trained with Different kind of features

Chapter 4

Optimized Classifier

4.1 Introduction

ANN structure trapped into local minima when input data is redundant. Redundancy within the features and irrelevant features in terms of classification task increase modeling complexity without improving discrimination performance [23], [24], [25]. Since, there is no unanimity in selection of features for characterizing EEG signals, a scheme that in the first place includes all the reasonable features and then removing irrelevant may prove to be useful. The difficulty of ANN is its empirically chosen structure of the network i.e., the number of hidden units and their interconnections are chosen arbitrarily. There is no appropriate method to decide the necessary structure for a given application or size of training set [22]. Number of neurons in the hidden layer of the neural network has significant effect on the performance of network [26], [27]. More neurons in the hidden layer require more computation. Less number of neurons gives high training error and high generalization error due to under fitting. The performance decreases if solution traps in local minima for multi layer perceptron(MLP). In this work, genetic algorithm (GA) is used to optimize the committee neural network since it has the following advantages: (i) GA is a global optimization method based on the natural selection and genetic theory. GA can help to find out the optimal solution globally over a domain [28]- [29]. (ii) GA begins randomly generated population of the initial solution. Each candidate (chromosome) of population represents the one solution of the given problem. New solution is generated by the reproduction of the initial solution. Reproduction combines the three operations: Selection, Crossover and Mutation. Here, the classifier is optimized sequentially i.e. first the unnecessary input nodes are pruned followed by deletion of hidden nodes from different ANN structure used.

4.2 Genetic Algorithm

GA is a global optimization method based on the natural selection and genetic theory [28], [29]. GA performs exceptionally well at obtaining the global solution when optimizing difficult non-linear functions [31]. It has been applied in different areas such as fuzzy control [32], path planning [33], greenhouse climate control [34], modeling and classification [35]. GA searches from one population of solutions to another focusing area for the best solution so far, while continuously sampling the total parameter space. In general, GA [36] begins with randomly generated population of initial solutions. Each solution is called as a chromosome.

This population represents the first generation from which genetic algorithm begins its search to find an optimal solution. Genetic algorithm searches the solution of a given problem in a multi direction simultaneously. This enhances the probability of finding the global solution. Each chromosome (solution) is evaluated on the basis of a fitness function. Next generation of chromosome is obtained by randomly selecting the chromosome from the former generation. The algorithm of GA is given in algorithm 1. Encoding is the first

Algorithm 1 Genetic Algorithm

```

1:  $t \leftarrow 0$ 
2:  $Initipopulation \leftarrow p(t)$ 
3: while  $t \leq max_{generation}$  do
4:    $Evaluation \leftarrow Eava(p(t))$ 
5:    $p'(t) \leftarrow Selection(p(t))$ 
6:    $ch1(t) \leftarrow Crossover(p'(t))$ 
7:    $ch2(t) \leftarrow Mutation(ch1(t))$ 
8:    $p(t) \leftarrow Replacement(ch2(t))$ 
9:    $t \leftarrow t + 1$ 
10: end while

```

step of GA. All the parameters of a problem are encoded as an artificial chromosome for optimization. Chromosome is a string of genes with fixed length.

- *Intipopulation*: A random population of N chromosome is created. Each member of population which is known as a chromosome, represents the initial solution of the given problem.
- *Evaluation*: Chromosome is evaluated on the basis of fitness function f .
- *Selection*: This operation is used to select the chromosomes from current generation $p(t)$ for reproduction. Roulette wheel selecting technique which copies the N chromosomes from $p(t)$ into an intermediate population $p'(t)$ is used for this purpose. The chromosome having high fitness value has more chance to select for the reproduction.
- *Crossover*: In this operation some pair of selected chromosomes from the $p'(t)$ interchange their bit values after randomly generated bit positions with probability P_c . P_c = probability of crossover. After this operation intermediate population change to child population $ch1(t)$.
- *Mutation*: It is bitwise operation. In this operation some bits of the selected chromosomes from the $ch1(t)$ are toggled with the probability of p_m and produce the population of child chromosomes $ch2(t)$. P_m = probability of mutation.
- *Replacement*: In this operation initial population $p(t)$ is replaced by the $ch2(t)$. This represents the second generation. *Evaluation* to *Replacement* process is continued up to

maximum generation.

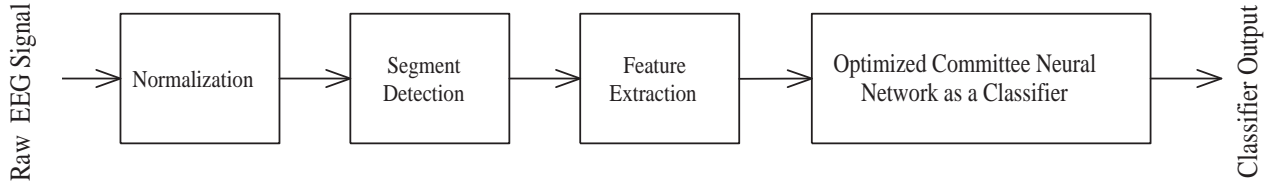


Figure 4.1: Flow diagram for classification of EEG signal

4.2.1 Feature selection and Hidden node optimization

Feature Selection

Feature selection is a task of selecting those features that are more useful for a classification problem from all those available. The purpose of feature selection is to reduce the input features length used in classification while maintaining acceptable classification accuracy. Redundant features are eliminated, leaving a subset of the original features having sufficient information to discriminate well among classes. IN pattern recognition techniques, the patterns are represented as a feature vector. The performance of the resulting classification algorithm are affected by the selection of features.

Consider a feature set, $F = (f_0, f_1, \dots, f_N)$. If f_0 and f_1 are dependent, then one of these could be discarded and the classifier has no less information to work with it. This has the benefit that computational complexity is reduced as there is smaller number of inputs and another benefit is that the accuracy of the classifier increases. This implies that the removed features were not adding any useful information but they were also actively hindering the recognition process. In this work we have used GA method for feature selection. Direct encoding of binary vector representation has been used to represent the chromosome. For feature selection i^{th} feature is selected if the i^{th} bit of corresponding chromosome is 1. If this bit is 0 then corresponding feature is not selected as shown in the Fig. 4.2

Hidden node Optimization of ANN

Artificial Neural networks is a successful tool in the fields of classification, prediction, pattern recognition etc. The hardware or software modeling of ANN has two distinct steps

1. Choosing proper network architecture.

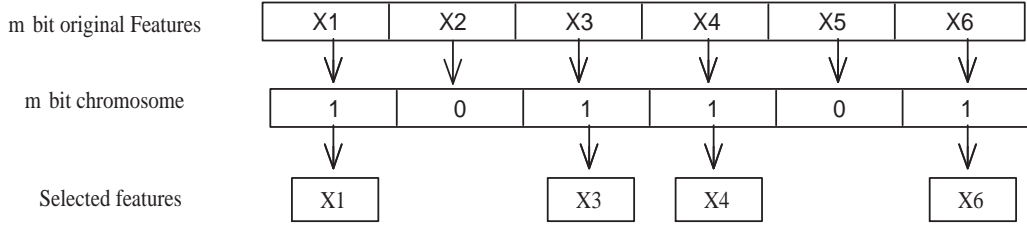


Figure 4.2: Representation of chromosome for feature Selection

2. Adjusting the parameters of a network so as to minimize certain fit criterion.

Performance of ANN is affected by the Size of Hidden node. Less number of hidden nodes generally result in under-fitting problem. On the other hand if the hidden nodes are too large the network will suffer from over-fitting problem where it can only make good prediction during training phase but poor performance in testing phase [44]. If the complexity of the problem is unknown the network architecture is set arbitrarily or by trial and error too small networks cannot learn the problem well, but too large network size leads to over fitting and poor generalization performance. So accordingly a network has to be evolved using evolutionary algorithms. For this purpose genetic algorithm is used to find the optimum number of neuron in the hidden layer.

Many techniques has been used to optimizing the hidden node. These techniques can be divided in to three categories. In first category network is generated with the small number of hidden nodes and it is increased until the maximum accuracy is reached. This method is called the constructive method. While second group of method includes the generation of network with the large number of hidden nodes initially and it is gradually decreases until the maximum accuracy is reached this is called the destructive method. Last one is genetic algorithm (GA). Optimization of hidden node using GA, indirect encoding has been used. Decoding of chromosome is done as given in Fig. 4.3

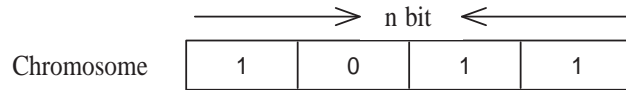


Figure 4.3: Representation of chromosome for hidden node encoding

$$Decimal_{value} = f(1)2^0 + f(2)2^1 + \dots + f(n)2^{n-1} \quad (4.1)$$

$$Decoded_{value} = 2 + (N_{max} - 2) \frac{Decimal_{value}}{(2^n - 1)} \quad (4.2)$$

Where N_{max} = maximum number of hidden nodes, and $2 < decode_value < N_{max}$.

Feature selection and hidden node optimization is done by GA that uses binary chromosome representation. The chromosome that represent the 0, 1 are divided in two parts. First part (m bit) is used for the input feature selection and second part (n bit) is used for the hidden node. As shown in the Fig. 4.4. First m bit of chromosome which is used for feature selection encoded by direct encoding scheme and the last n bit of chromosome which is used for hidden node uses indirect encoding scheme. This encoding scheme is called hybrid encoding scheme.

Algorithm for feature selection and hidden node optimization is as flow:

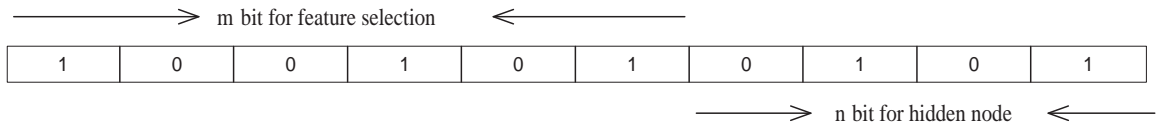


Figure 4.4: Representation of chromosome for feature selection and hidden node optimization

Step 1: Generate the initial population $X(0) = [x_1(0), x_2(0), x_3(0), \dots, x_N(0)]$, each $x_i(0)$ represent an encoded network at $t = 0$. N represents the size of population. Here hybrid encoding is used which is combination of direct and indirect encoding scheme, as discussed above direct encoding is used for feature selection and indirect encoding for hidden node. Decode the each individuals (chromosome) in to a set of hidden neuron and selected input feature. Construct the fully connected ANN from each decoded chromosome. So N number of ANN are constructed according to the above encoding method.

Step 2: Apply the corresponding training data to each network. Train the each network. After training, apply the corresponding testing data to each network and compute the fitness of each individual (chromosome).

Fitness calculation: Choice of fitness function is important factor for a given problem because it uniquely differentiate the one chromosome from the other chromosome. Therefore we can say fitness function characterize the chromosome. To calculate the fitness value of each chromosome, network is trained with the training data with some predefined fixed value of training error as known that the low error value in the training phase does not represent the sufficient condition for suitable network performance during testing phase [46]. Taking into account this condition testing phase is performed with the corresponding testing data and testing accuracy is calculated. In present work we are optimizing both the hidden node

and input feature. Number of input feature which is not selected is consider in the fitness function. Hence our Fitness function depend on the two factors, one is testing accuracy that is out of total classes how many number of classes it correctly classified and Second is out of total feature how many feature is not selected. Combining these two term fitness function is given as:

$$Fitness = A * \frac{Correctly}{TP} + B * \frac{Zeros}{TF} \quad (4.3)$$

Where *Correctly* = Total number of correctly classified patterns, *TP* = Total number of testing patterns,

Zeros = Total number of features minus selected number of features, and *TF* = Total number of features. Multiplication of *Correctly* with higher value of *A* (10^4) and *Zeros* with lower value of *B* (0.4).

Step 3: Select parents chromosomes from the current generation based on their fitness value. Chromosome having highest fitness value have more chance to be selected. Number of selected parents is *N*. Selection is based on the roult wheel selection.

Step 4: Apply the crossover operator to the parents to generate the offspring (child chromosome). This child chromosome will replace only those parent chromosome which has participated in the crossover operation. Now mutation is applied to the parent chromosomes. Generated childs by mutation operation replaces selected parents which has participated in the mutation operation, this forms the next generation and set $t = t + 1$.

Step 5: Decide whether termination criteria are reached or not. If yes stop otherwise go to step 2. Stopping criterion is iteration number (maximum number of generation). Fig. 4.5 represent the above algorithm.

4.2.2 Experiment Using ANN and Committee Neural network with selected feature and optimized hidden neurons

Here feature vector is mixture of features (31 dimension). CNN and ANN are optimized by selecting the input features from 31 features and selecting the number of neurons in the hidden layer from initial number of neuron. CNN is optimized by optimizing its member neural networks. Initial population of binary chromosome is generated for each network. Length of chromosome for each network is different because the initial number of hidden neurons for each network is different. Population size for each network is different (30, 40, and 35) and the maximum number of generation (iteration number) is 70. Probability of

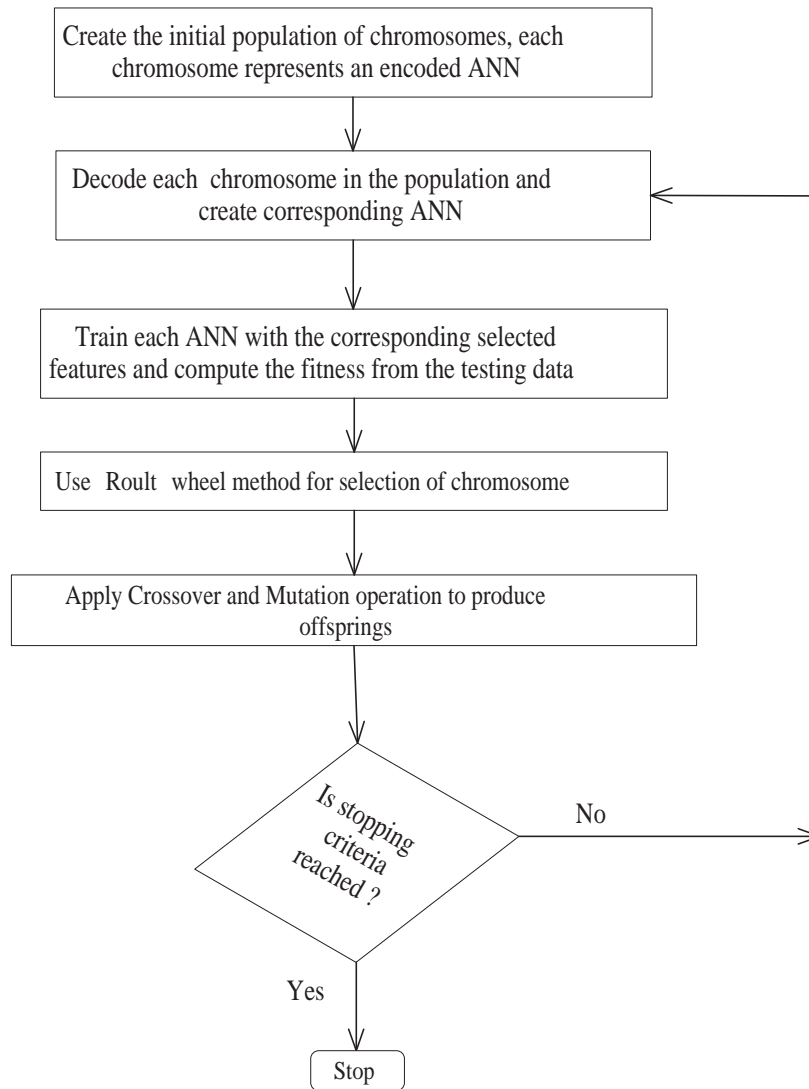


Figure 4.5: Flow chart of feature selection and hidden node optimization

crossover (P_c) and probability of mutation is 0.7 and 0.01 respectively for each network. At each generation we have to train the 40 networks so for 50 generation total number of trained network is 2000, it takes large training time. In order to reduce the training phase time the network was not trained up to maximum iteration to get the desired less training error, instead of that the network was trained up to some fixed number of less iterations (user defined). If the network achieves the desired training error in that given iteration then it will stop. If network does not achieve the desired training error in that given iteration then it again trained up to some other iteration and the change in the mean square error is calculated. If this change is greater than the 0.0001 then the network is again trained this training is continue until the change in the error is not less than 0.0001.

when the maximum number of generation is reached , the maximum fittest chromosome from the whole generation is selected as the optimized chromosome and corresponding decoded value gives the selected feature and number of neurons in the hidden layer. These selected features of training data are applied to the input layer of network and training is performed. Stopping criteria of neural network is mean square error. After training, selected features of the testing data corresponding to optimize chromosome are applied to the neural network and the confusion matrix of the classified data is given in the table below. Classification accuracy of all the neural networks is shown in the Table 4.5. From Table 4.5 it is clear that all the neural networks have gained the accuracy with the use of genetic algorithm. Maximum accuracy is obtained is 96.5 percent corresponding to ANN which is more than previous accuracy result. Out of different neural networks accuracy of ANN is more because it uses more number of training data set. Output of these optimized hidden neuron network

Classifier	Desired Result	Out put		
		SET A	SET D	SET E
ANN	SET A	793	24	00
	SET D	07	737	14
	SET E	00	39	786

Table 4.1: Confusion matrix of ANN with selected features and optimized hidden neurons with selected input feature are used to form the committee neural network as shown in the Fig. 3.8. The confusion matrix of the testing data are shown below in the Table 4.3 for level 1 and level 2. Classification accuracy of the Committee neural network at level-1 and level-2 are shown below in the Table 4.5. Reduced feature length and the number of optimized hidden neurons for each network, and the original full feature length with initial number of hidden neurons in each network are shown below in the Table 4.4. Proposed technique is compared for the same data base with standard artificial neural network with MLP-BP algorithm, combined neural network model based [12] and wavelet / mixture of experts network based [11] approach. Comparison has been shown in the Table 4.6. From Table 4.6, It is verified that the proposed method provides the more accurate results than the wavelet/mixture of experts network based and Combined neural network based model.

Classifier	Desired Result	Out put		
		SET A	SET D	SET E
NN1	SET A	786	38	67
	SET D	13	741	41
	SET E	1	21	692
NN2	SET A	764	39	13
	SET D	29	729	26
	SET E	07	32	761
NN3	SET A	783	52	19
	SET D	06	688	06
	SET E	11	60	775
NW1	SET A	776	30	00
	SET D	22	754	27
	SET E	02	16	773
NW2	SET A	776	31	32
	SET D	23	723	09
	SET E	01	46	759

Table 4.2: Confusion matrix of NN1, NN2, NN3, NW1, and NW2 with selected features and optimized hidden neurons

Classifier	Desired Result	Out put		
		SET A	SET D	SET E
CNN-1	SET A	793	28	22
	SET D	03	737	19
	SET E	04	35	759
CNN-2	SET A	794	18	02
	SET D	06	759	11
	SET E	00	23	787

Table 4.3: Confusion matrix of CNN-1 and CNN-2 with selected features and optimized hidden neurons

Neural Network	Original Feature Length	Selected Feature Length	Number of Hidden Neurons	Optimized Hidden Neurons
ANN	31	23	11	10
NW1	31	19	11	10
NW2	31	19	11	10
NN1	31	19	7	5
NN2	31	18	22	10
NN3	31	19	33	15

Table 4.4: Selected features length and optimized hidden neurons

Classifier	Accuracy(In percent) with full features	Accuracy(In percent) with selected features
ANN	95.96	96.50
NW1	95.04	95.95
NW2	93.00	94.08
NN1	91.95	92.45
NN2	93.58	93.91
NN3	93.08	93.58
CNN-1	96.04	95.375
CNN-2	97.29	97.50

Table 4.5: Accuracy matrix of different classifiers with full features and selected features

Statistical parameter(in %)	wavelet/mixture of experts network based model	Combined neural network based model	Proposed model	
			CNN-2	Optimized CNN-2
Specificity	94	96	99.25	99.25
Sensitivity	92.5	94.5	94.75	94.875
Sensitivity	93	94	97.875	98.375
Accuracy	93.17	94.8	97.29	97.5

Table 4.6: Comparison of proposed method with wavelet/mixture of experts and Combined neural network based model

Chapter 5

Concluding Remarks

5.1 Conclusion

Electroencephalography (EEG) signals provide valuable information to study the brain function and neurobiological disorders. Digital signal processing gives the important tools for the analysis of EEG signals. The EEG signal was collected from the standard data base. These EEG signals were not distinguishable with human eyes. We used the signal processing tools to distinct them and Provide the status of the individual. DWT and AR method are used for the feature extraction. Performance of proposed mixture of features is compared to the DWT features and AR features using ANN classifier. Experiment result shows that the mixture of features classify the EEG signals more accurately. Therefore, the two feature extraction schemes are partially complementary in nature whose combination helps to represent and characterize the signal more effectively than using any of them alone. Performance of proposed Two-level committee neural network shows that the accuracy increases from 96.04 percent to 97.29 percent. Therefore two-level CNN classify the three data set more accurately than single MLP-BP network. Genetic algorithm is used to optimized the each and every member network of committee neural network. Each optimized member network of CNN intern optimized the CNN. Experiment result shows that the every optimized member network of CNN classify the EEG signal more accurately than the unoptimized network. Therefore performance of the optimized CNN is increased and the structural complexity of network is reduced. Accuracy of optimized CNN is increased from 97.29 percent to 97.5 percent. Result of the proposed classifier is compared with the result of combined neural network model based [12] and wavelet / mixture of experts network based [11] approach in terms of statistical parameters. Comperission result shows that the performance of CNN classifier is superior. It is noticed that the performance of the proposed technique is superior than each of the above established technique. Therefore CNN classifier can be used to classify the EEG signals more accurately.

5.2 Future work

Looking for further improvement in accuracy by combination of features extracted by Lyapunov Exponents and wavelet transform. Use optimization algorithm in Lyapunov Exponents feature extraction based methods and support vector machine as a classifier (svm).

Chapter 6

References

Bibliography

- [1] E. D. Übeyli, “Statistics over features: EEG signals analysis,” *Computers in Biology and Medicine*, vol. 39, Issue 8, pp. 733-741, 2009.
- [2] Ali B. Usakli, “Improvement of EEG Signal Acquisition: An Electrical Aspect for State of the Art of Front End,” *Computational Intelligence and Neuroscience*, vol. 2010, 2009.
- [3] E. D. Übeyli, “Measuring saliency of features representing EEG signals using signal-to-noise ratios,” *Expert Systems with Applications*, vol. 36, pp. 501-509, 2009.
- [4] Saeid Sanei and J.A. Chambers, *EEG Signal Processing*, John Wiley and Sons Ltd, England, 2007
- [5] H.L. Attwood and W. A. MacKay, *Essentials of Neurophysiology*, B. C. Decker, Hamilton, Canada, 1989
- [6] M. Teplan, “Fundamentals of EEG measurements,” *Measmt Sci. Rev.*, vol. 2, no. 2, 2002.
- [7] <http://www.everydayhealth.com/health-center/electroencephalogram-eeeg-results.aspx>
- [8] H. Adeli ^a, Z. Zhou ^b, and N. Dadmehr ^c, “ Analysis of EEG records in an epileptic patient using wavelet transform,” *Journal of Neuroscience Methods*, vol. 123, pp. 69-87, 2003.
- [9] M. Unser and A. Aldroubi, “A review of wavelets in biomedical applications,” *Proceedings of the IEEE*, vol. 84, Issue 4, pp. 626-638, 1996.
- [10] N. Hazarika, J. Z . Chen, and A. Chung Tsoi, “Classification of EEG signals using the wavelet transform,” *Signal Processing*, vol. 59, pp. 61-72, 1997.
- [11] E. D. Übeyli, “Wavelet/mixture of experts network structure for EEG signals classification,” *Expert Systems with Applications*, vol. 34, pp. 1954-1962, 2008.

- [12] E. D. Übeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification," *Digital Signal Processing*, vol. 19, pp. 297-308, 2009.
- [13] E. D. Übeyli, "Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals," *Expert Systems with Applications*, vol. 37, pp. 233-239, 2010.
- [14] Steven M. Kay and S. Lawrence Marple, "Spectrum Analysis-A Modern Perspective," *Proceedings of the IEEE*, vol. 69, no. 11, 1981.
- [15] P. Stoica and R. L. Moses, *Introduction to spectral Analysis*, Upper Saddle River, NJ: Prentice-Hall, 1997.
- [16] J. P. Burg, "Maximum entropy spectral analysis," *37th Meeting Society of Exploration Geophysicists*, Oklahoma City, OKla, 1967.
- [17] <http://www.dtic.upf.edu/xserra/cursos/TDP/referencies/Park-DWT.pdf>.
- [18] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 626638, 1996.
- [19] S. Haykin, *Neural Networks: A Comprehensive Foundation*, New York: Macmillan, 1994.
- [20] A. J. C. Sharkey, "On combining artificial neural nets," *Connection Science*, vol. 8, no. 3/4, pp. 299-314, 1996.
- [21] N. P. Reddy and O. A. Buch, "Speaker verification using committee neural networks," *Computer Methods and Programs in Biomedicine*, vol. 72, Issue 2, pp. 109-115, 2003.
- [22] S. Ari and G. Saha, "In search of an optimization technique for Artificial Neural Network to classify abnormal heart sounds," *Applied Soft Computing*, vol. 9, Issue 1, pp. 109-115, 2009.
- [23] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality Reduction Using Genetic Algorithms," *IEEE transactions on evolutionary computation*, vol. 4, no. 2, 2000.

- [24] Jihoon Yang and Vasant Honavar, "Feature Subset Selection Using a Genetic Algorithm," *Genetic Programming 1997: Proc. of the Second Annual Conference*, vol. 13, 1997.
- [25] Dash M and Liu H, "Feature selection methods for classifications," *Intell Data Anal*, vol. 1(3), pp. 131-156, 1997.
- [26] W. Wang, W. Lu, A. Y. T. Leung, Siu-Ming Lo, Z. Xu, and X. Wang, "Optimal Feed-Forward Neural Networks Based on the Combination of Constructing and Pruning by Genetic Algorithms," in *Proc. of International Joint Conference*, pp. 636-641, 2002.
- [27] L. Fletcher, V. Katkovnik, and F. E. Steffens, "Optimizing the number of hidden nodes of a feed-forward artificial neural network," *The 1998 IEEE International Joint Conference*, pp. 1608-1612, 1998.
- [28] J. H. Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor: The U. of Michigan Press, 1975.
- [29] D. T. Pham and D. Karaboga, *Intelligent Optimization Techniques, Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*, New York: Springer-Verlag, 2000.
- [30] <http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html>
- [31] F. H. F. Leung, H. K. Lam, S. H. Ling, and P. K. S. Tam, "Tuning of the structure and parameters of a neural network using an improved genetic algorithm," *IEEE Trans. on Neural Networks*, vol. 14, pp. 79-88, 2003.
- [32] H. K. Lam, S. H. Ling, F. H. F. Leung, and P. K. S. Tam, "Tuning of the structure and parameters of neural network using an improved genetic algorithm," in *Proc. 27th Annu. Conf. IEEE Ind. Electron. Soc.*, Nov. 2001, pp. 25-30.
- [33] G. P. Miller, P. M. Todd, and S. U. Hegde, "Designing neural networks using genetic algorithms," in *Proc. 3rd Int. Conf. Genetic Algorithms Applications*, pp. 379-384, 1989.
- [34] N. Weymaere and J. Martens, "On the initialization and optimization of multilayer perceptrons," *IEEE Trans. Neural Networks*, vol. 5, pp. 738-751, 1994.
- [35] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

- [36] Zbigniew and Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, London, Springer-Verlag, 1996.
- [37] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Information Theory*, vol. 36, no.5, pp. 961-1005, 1990.
- [38] S. Soltani, "On the use of the wavelet decomposition for time series prediction," *Neuro-computing*, vol. 48, pp. 267-277, 2002.
- [39] Petra Nass. (1999) The Wavelet Transform. [Online].
Available: <http://www.eso.org/sci/software/esomidass/doc/user/98NOV/volb/node308.html>
- [40] R. Q. Qüiroga and M. Schurmann, "Functions and sources of event-related EEG alpha oscillations studied with the Wavelet Transform," *Clinical Neurophysiology*, vol. 110, pp. 643-654, 1999.
- [41] Lori Mann Bruce, Senior Member, Cliff H. Koger, and Jiang Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature fxtraction," *IEEE Trans. On Geoscience and Remote Sensing*, vol. 40, no. 10, 2002.
- [42] C. Valens. (1999) A Really Friendly Guide to Wavelets. [Online].
Available: <http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html>
- [43] I. Gler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, Issue 2, pp. 113-121, 2005.
- [44] S. Lawrence, C. Giles, and A. Tsoi, "Lessons in neural network training: Over fitting may be harder than expected," *In Proc. of the Fourteenth National Conference on Artificial Intelligence*, pp. 540 -545, 1997.
- [45] Yung-hsiang Chen ^{a,b} and Fi-John Chang ^a, "Evolutionary artificial neural networks for hydrological systems forecasting," *Journal of Hydrology*, vol. 367, pp. 125137, 2009.
- [46] P. Arena, R. Caponetto, L. Fortuna, and M. G. Xibilia, "genetic algorithms to select optimal neural network topology," *In Proc. of the 35th Midwest Symposium*, vol. 2, pp. 1381-1383, 1992.