# P300 DETECTION FOR BRAIN COMPUTER INTERFACE

Thesis submitted in partial fulfilment

for the award of the degree of

**Master of Technology**

*in*

**Electronics and Instrumentation Engineering**

*by*

**Deepesh Kumar**

**211EC3312**

**Department of Electronics & Communication Engineering**

**National Institute of Technology, Rourkela-769 008, India**

**2013**

# P300 DETECTION FOR BRAIN COMPUTER INTERFACE

Thesis submitted in partial fulfilment

for the award of the degree of

**Master of Technology**

*in*

Electronics and Instrumentation Engineering

*by*

**Deepesh Kumar**

**211EC3312**

*under the guidance of*

**Prof. Samit Ari**

**Department of Electronics & Communication Engineering**

**National Institute of Technology, Rourkela-769 008, India**

**2013**

# Declaration

I hereby declare that the work presented in the thesis entitled "*P300 detection for brain computer interface"* is a bona fide record of the research work done by me under the supervision of Prof. Samit Ari, Department of Electronics & Communication Engineering, National Institute of Technology, Rourkela, India and that no part thereof has been presented for the award of any other degree.

**Deepesh Kumar**
Dept. of Electronics & Comm. Engg.
National Institute of Technology
Rourkela, India-769 008

*dedicated to*

*my parents with love..*

# National Institute of Technology
# Rourkela

## *Certificate*

This is to certify that the work in the thesis entitled “P300 Detection for Brain Computer Interface” by Deepesh Kumar is a record of an original research work carried out under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of Master of Technology in Electronics and Communication Engineering. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

**Prof. Samit Ari**

Dept. of Electronics & Comm. Engg.

National Institute of Technology, Rourkela

Rourkela, India-769 008

# Acknowledgement

It is my pleasure and privilege to thank many individuals who made this report possible. First and foremost I offer my sincere gratitude towards my supervisor, Prof. Samit Ari, who has guided me through this work with his patience. A gentleman personified, in true form and spirit, I consider it to be my good fortune to have been associated with him. I thank Prof. Sukadev Meher, Head, Dept. of Electronics & Comm. Engg., NIT Rourkela, for his constant support during the whole semester.

I would like to mention the names of Manab, Dipak, Girish, 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

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.

**Deepesh Kumar**

**(211EC3312)**

# Abstract

P300 based brain computer interface (BCI) sometimes called brain machine interface (BMI) is a way of direct communication between human brain and external device which provides an alternative communication link with outside world to the people who are unable to communicate via conventional means because of sever motor disability. P300 wave is an event related potential which evoked in the process of decision making of human brain which can be generated using oddball paradigm. This thesis aims to detect the P300 wave as accurate as possible. To do that this study proposed discrete wavelet transforms (DWT) based feature extraction method from each P300 and No-P300 of EEG signal from the entire 64 channel. Principal component analysis (PCA) technique is further applied for the reduction of the dimension of the feature. Detection of P300 is achieved using support vector machine (SVM) and artificial neural network (ANN) classifier. Experimental result shows that the proposed method with SVM classifier yields better performance compared to the method with ANN.

*Keywords:* Artificial Neural Network (ANN), Brain Computer Interface (BCI), Discrete Wavelet Transform (DWT), Electroencephalography (EEG), Event Related Potential (ERP), P300, Support Vector Machine (SVM).

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Brain Computer Interface System

Any natural form of communication or control requires peripheral nerves and muscles. It starts with the user's intent. This intent triggers a process in which certain brain areas are activated, and hence signals are sent via the peripheral nervous system (specifically, the motor pathways) to the corresponding muscles, which in turn perform the movement necessary for the communication or control task. The activity resulting from this process is often called motor out-put. A brain computer interface (BCI) provides an alternative way to the natural communication and control. A BCI is an artificial system that bypasses the body's normal efferent pathways, which are the neuromuscular output channels. Instead of depending on peripheral nerves and muscles, a BCI directly measures brain activity associated with the user's intent and translates the recorded brain activity into corresponding control signals for BCI applications. This translation involves signal processing and pattern recognition, which is typically done by a computer. Since the measured activity originates directly from the brain and not from the peripheral systems or muscles, the system is called a Brain–Computer Interface. A BCI must record brain activity directly from scalp invasively or non-invasively and it should provide feedback to the user in real time and also it must rely on international control. [1]. Figure 1.1 is considered as the main identifiers of a BCI system and used by many BCI researchers [2], [3], [4].

Brain-Computer interface (BCI) provides an alternative communication between a human brain and an external device for the peoples who are unable to communicate via conventional means because of severe motor disabilities like spinal cord injuries [1]. For BCI the brain signals are acquired by electroencephalography (EEG) which is a type of non-invasive method for monitoring brain activity. EEG signal provides visual display of the recorded waveform and allows computer aided signal processing techniques to characterize them which further enable us to apply the advanced digital signal processing techniques for analysis of EEG signals [5].

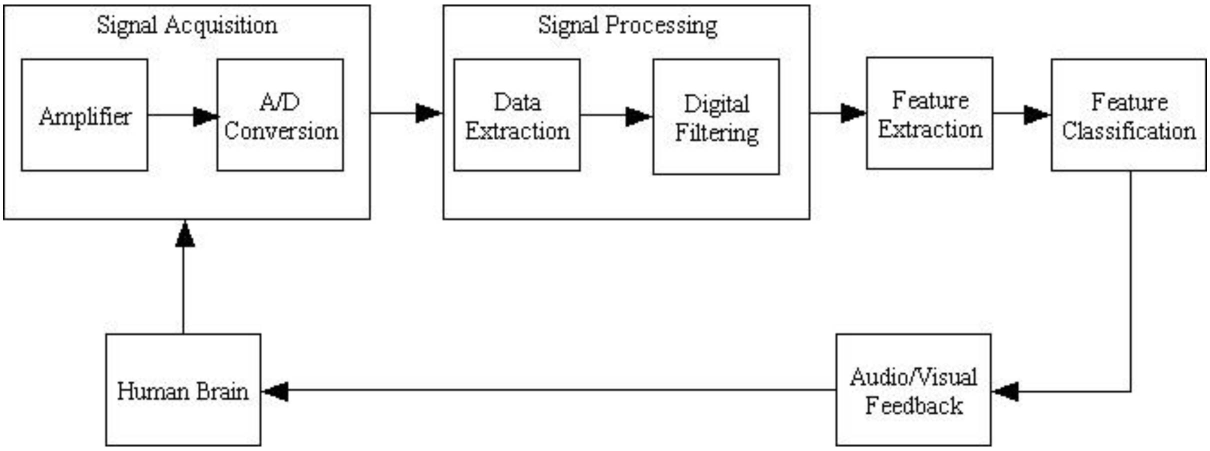


Figure 1.1: Basic design and operation of any BCI system

Over the last two decades, there have been numerous studies performed on BCI. Researchers proposed various methodologies, extended the application fields of BCI and investigated the physiological nature of the experimental paradigms [1]. However, the main challenge in BCI is to improve the usability and practicality of these systems. Thus, researchers put most of their effort on developing new algorithms to improve the speed and accuracy of the prediction mechanisms in BCI applications. Due to the nature of existing BCI's, the applications are considered as pattern recognition problems and variety of signal processing, feature extraction and pattern classification techniques are being experimented in these systems [6].

This thesis deals to one area of BCI which is based on P300 wave detection through spelling paradigm (P300 speller). First introduced by Farwell and Donchin in 1988 [7], Spelling Paradigm enables paralyzed people to express their thoughts and feelings by spelling the words on a computer screen. In this application it is aimed to detect P300 correctly so as to that the subject thinks of by presenting some visual stimuli to the subject. In order to accomplish this task, the brain activity is acquired by using bio potential measurement devices (usually via Electroencephalography - EEG) and further analysed with advanced signal processing and classification methods [6].

## 1.2 Brain Signal Measurement

Basically electroencephalography (EEG) is used to measure the brain activity from the scalp in the form of electrical signal. The electroencephalogram (EEG) is defined as electrical activity of an alternating type recorded from the scalp surface after being picked up by metal electrodes and conductive media [1]. The EEG signal is nothing but the current measurement between the dendrites of nerve cells in the cerebrum region of the brain during their synaptic excitation. This current consists of electric field measured by electroencephalography (EEG) equipment and the magnetic field quantified by electromyogram (EMG) devices [8].

Our brain has three different region namely, cerebrum, cerebellum and brain stem. Each region of the brain represent different status of human body like the cerebellum region represent initiation of movement, consciousness and state of mind, the cerebellum region plays a role in voluntary action like muscle related movement and the brain stem region control the respiration functioning and neural hormones. So it is very much clear that the EEG signal generated from brain certainly express the status of whole body and brain disorder [9].

### *1.2.1 EEG Signal Recording*

In conventional scalp EEG, the recording is obtained from many electrodes arranged in a particular pattern on the scalp with a conductive gel or paste. Electrode locations and names are specified by the International 10–20 system for most clinical and research applications [10].

Measuring electrodes read the signal from the scalp which is in the range of microvolt thus amplifier is used to amplify the bio potential in to the range where they can be digitised accurately then A/D converter converts analog signal to digital signal and finally personal computer or relevant device stores and displays recorded data. Scalp recordings of neuronal activity in the brain, identified as the EEG, measures the potential changes over time in basic electric circuit conducting between signal (active) electrode and reference electrode. Extra third electrode, called the ground electrode, is required for getting differential voltage by subtracting the same voltages showing at active and reference points. The mono-channel EEG measurement consists of one active electrode, one (or two specially linked together) reference and one ground electrode. The multi-channel configurations may comprise up to 128 or 256 active electrodes.

### *1.2.2 Brain Waves Classification*

Most of the EEG waves lies in the range of 0.5 to 500 Hz, however there are 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 with highest amplitude. It is dominant rhythm in infants up to one year and also in adults during deep sleep.

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

**Alpha waves:** Alpha waves have frequency range from 7 HZ to 12 Hz. it is seen most commonly in adults. Alpha activity occurs 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 but having small amplitude. It lies in the frequency range from 14 HZ to 30 Hz. It is the dominant rhythm in the persons who are alert 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. The amplitude of beta wave is less than 30 µV.

**Gamma waves:** It is fastest waves of brain having frequency range from 30-45 Hz with very low amplitude; however the detection of these rhythms plays an important role in finding the neurological diseases. These waves generally occured in front central part of the brain. It suggests the event-related synchronization (ERS) of the brain [8].

## 1.3 Event Related Potential

Event Related Potentials are specific patterns, occurs when an auditory or visual stimulus is presented. These include the P300 patterns and Steady State Visual Evoked Potentials (SSVEP).

### *1.3.1 P300 signals*

The P300 (P3) wave is an event related potential (ERP) component elicited in the process of decision making. It is considered to be an endogenous potential, as its occurrence links not to the physical attributes of a stimulus, but to a person's reaction to it.

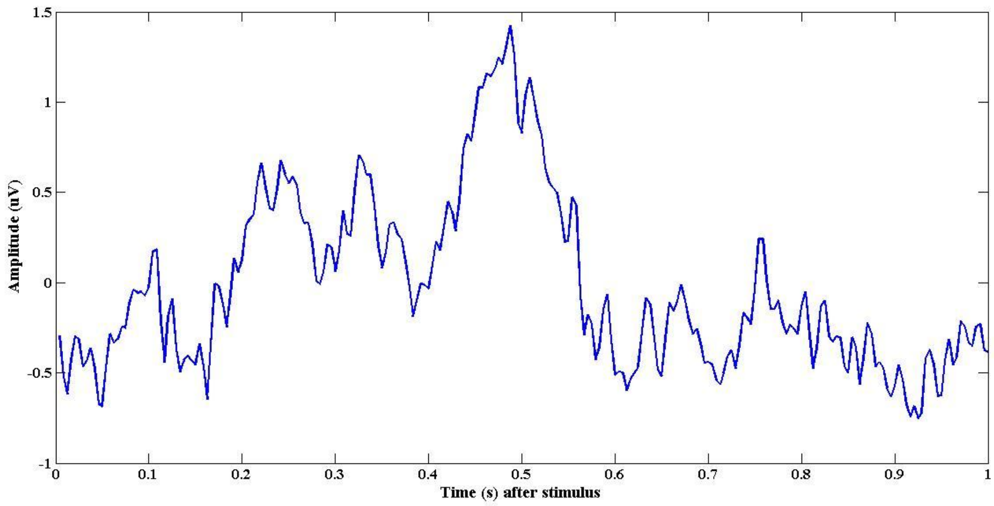


Figure 1.2: A typical P300 signal.

The wave corresponds to positive deflection in voltage at latency of about 300 ms in the EEG [11]. In other words, it means that after an event like a flashing light, a deflection in the signal should occur after 300ms. Furthermore, the presence, magnitude, topography, and time of this signal are often used as metrics of cognitive function in decision making processes. If P300 wave is detected 300 ms after a flashing light in a specific location, it means that the user was paying attention to this same location. The detection of a P300 wave is equivalent to the detection of where the user was looking 300 ms before its detection. Signal is typically measured most strongly by the electrodes covering the parietal lobe. However, Krusienski *et al*. [12] showed that occipital sites are more important In a P300 speller; the main goal is to detect the P300 peaks in the EEG accurately and instantly. The accuracy of this detection will ensure a high information transfer rate between the user and the machine. [7].

### *1.3.2 Steady State Visual Evoked Potentials*

Steady State Visual Evoked Potentials (SSVEP) are oscillating signal patterns elicited in the brain according to the frequency of the presented periodic visual stimulation. These signals are more distinctive in occipital regions of the brain which is believed to be related to visual activities. SSVEP is employed in BCI applications by the presentation of several flickering light sources with different frequencies [1]

## 1.4 Literature Survey

Ben H. Jansen *et al.* proposed a threshold detector for single-trial P300 detection which operates on the 0–4 Hz band, isolated from the raw electroencephalogram using low-pass filtering, wavelet transforms, or the piecewise prony method (PPM). A detection rate around 70% was found, irregardless of stimulus type, inter-stimulus interval (ISI), probability of occurrence of the target stimuli, intra session and intersession effects, or filtering method [14]

Min Ki Kim *et al*. [15] and Rosas-Cholula [16] presented wavelet analysis and PCA pre-processing for detection of P300 rhythm using 14 channels which is further applied to input to classifier.

Matthias Kaper *et al.* proposed support vector machine approach to analyse the P300 speller paradigm. In this classification they found correct solution after five repetitions using only 10 electrode positions [17].

Hubert Cecotti *et al.* presented an approach of convolution neural network for detection of P300 waves. The topology of the network is adapted to the detection of P300 waves in the time domain. Seven classifiers based on the CNN are proposed: four single classifiers with different features set and three multi-classifiers.

Vladimir Bostanov presented the t-CWT method for feature extraction, a simple and fast CWT computation algorithm for the transformation of large data sets and single trials P300 detection [18].

Alain Rakotomamonjy and Vincent Guigue proposed ensemble of SVM approach for BCI P300 speller. They addresses the problem of signal responses variability within a single subject and copes with such variability's through an ensemble of classifiers approach.[19].

Kaper *et al*. investigated the application of Support Vector Machines (SVM's) using Gaussian kernel transformation but using only a few number of EEG channels [20]. They have been elected as one of the winners in BCI Competition II [21] because of the success of the method they have proposed. SVM has also been applied in numerous studies for the classification of P300 responses [22], [23] and in other BCI applications [24], [25], [26].

## 1.5 Objective

The main objective of our research is to correctly detect and classify the P300 and no P300 wave from the raw EEG signal acquired using P300 speller paradigm so as to make P300 based BCI system more reliable for practical application rather than just laboratory experiment. In order to achieve this following process is carried out in this thesis:

(i) EEG signal filtered and normalised within the pre-processing stage.

(ii) A feature extraction technique is proposed, which combines extracted coefficient of Daubechies-4 wavelet at second level decomposition,

(iii) To reduce the curse of dimensionality of feature a technique called principal component analysis (PCA) is proposed for feature reduction.

(iv) To resolve binary class classification problem of EEG data ANN and SVM classification technique is applied.

## 1.6 Data Base and Experimental Set-up

Data set II of the third BCI competition is used here for P300 detection [27]. This database was actually provided by the Wadsworth Centre, New York State Department of Health. The detail description of the database and the experimental setup they used for collecting the data is explained here.

### *1.6.1 Experimental Setup*

For complete recording of P300 evoked potential by BCI 2000, the P300 speller paradigm is used which is described by Dolchin *et al*, [28] and originally Farewell and Dolchin, 1988. Here user is presented with 6×6 matrix of alphanumeric character on the computer screen. User is asked to concentrate on specific characters in a word known as target character (one character at a time) which is predefined by the instructor. The rows and Columns of the matrix are intensified randomly with some specific duration and inter-stimulus interval which is here 5.7Hz and 75 ms respectively.

Here the row and column intensification constitute the visual stimulation for user's brain and also the number of target stimulation is very less compared to non target stimulation. For example, in the 6×6 matrix of characters, there is only 2 out of 12 intensification is target intensification, one for row and other for column, having desired character, rest of the 10 intensification constitutes the non-target intensification. The response of the user's mind to target and non-target intensification is supposed to be different. The underlying principle of this whole spelling paradigm is that the subject must produce specific response to the target intensification since it occur very less than non-target ones. The target intensification is expected to evoke the so called P300 potential which is nothing but a type of event related potential.

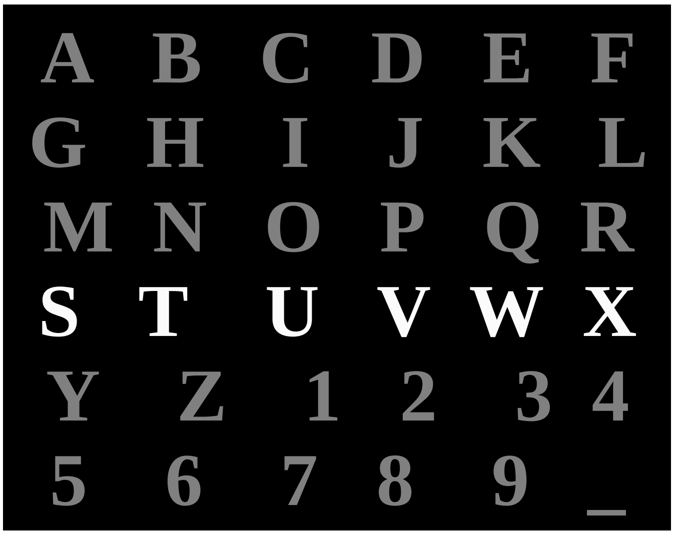


Figure 1.3: P300 Speller Matrix used in the datasets collection [29]

To correctly predict the target character it is very important that the target intensification stimulus response i.e. P300 wave should be clearly distinguishable from the non-target stimulus response i.e. no-P300. However usually it is very difficult to distinguish this in one trial i.e. by intensifying the row and column only ones. The reason behind this is that EEG signal are highly sensitive to noise and this makes it very difficult to distinguish the target response from the non-target response. Therefore, several trials are required for single target character in order to increase the prediction accuracy.

## *1.6.2 Data Collection Process*

Data set II of the third BCI competition gives the signals collected from two subjects (band pass filtered from 0.1-60 Hz and digitized at 240 Hz) in five sessions each. Each session comprises of a number of runs. In each run, the subject focused attention on a series of characters. For each character epoch in the run, user display was as follows: the matrix was displayed for a 2.5s period, and during this time each character had the same intensity (i.e., the matrix was blank). Subsequently, each row and column in the matrix was randomly intensified for 100ms (i.e., resulting in 12 different stimuli – 6 rows and 6 columns). After intensification of a row/column, the matrix was blank for 75ms. Row/column intensifications were block randomized in blocks of 12. The sets of 12 intensifications were repeated 15 times for each character epoch (i.e., any specific row/column was intensified 15times and thus there were 180 total intensifications for each character epoch). Each character epoch was followed by a 2.5s period, and during this time the matrix was blank. This period informed the user that this character was completed and to focus on the next character in the word that was displayed on the top of the screen (the current character was shown in parentheses).

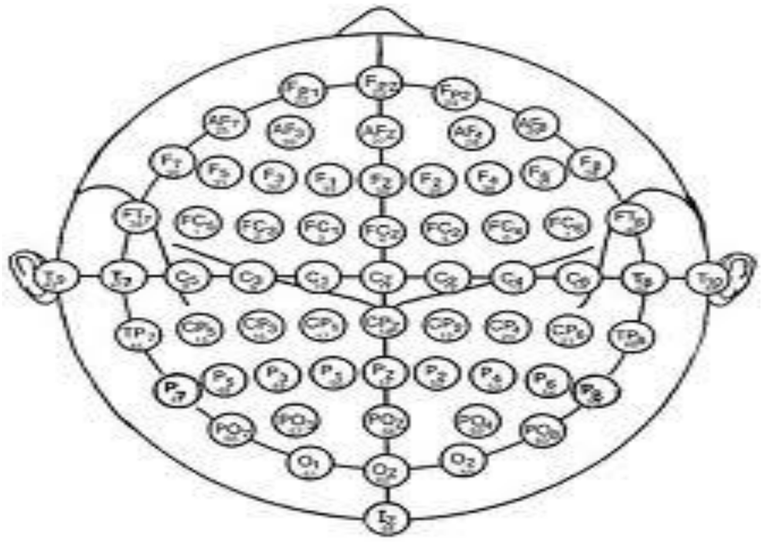

Figure 1.4: Electrode designations and channel assignment numbers for EEG measurement [28]

All of the rows/columns were intensified 15 times; therefore 30 possible P300 responses should be detected for single character recognition. Signals from the two subjects were collected from 64 ear-referenced channels as shown in Figure 1.4. The signal was band pass filtered from 0.1 to 60 Hz and digitized at 240 Hz. A more detailed description of the data-set can be found in the BCI competition page [27].

## 1.7 Thesis outline

**Chapter 1** of the thesis explain the background of P300 based BCI system and also give the information about EEG and literature survey to P300 detection with reference to BCI. Next it gives detail about the online data used in this thesis and the experimental setup to acquire the brain signal.

**Chapter 2** describe the pre-processing to the original EEG signal before feature extraction take place and then explain feature extraction using wavelet transform and also gives detail about feature reduction technique called PCA, this chapter also discuss result obtained from feature extraction and feature reduction.

**Chapter 3** explain the classification problem in detail and gives solution to this by using ANN and SVM classifier.

**Chapter 4** describes the classifier output which is final output of the whole thesis work and also gives the concluding remark and future work.

# Chapter 2

# Pre-processing and Feature Extraction

## 2.1 Introduction

The original biomedical signals collected from the subject are very large in dimension and also have redundant information which makes signal processing complex. To avoid this situation the original input data is transformed into reduced number of feature vector. Transforming the input data into set of feature is called feature extraction. The extracted feature set should give relevant information to the original signal without loss of important information. 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.

## 2.2 Wavelet Transform

A transform is the way of remapping the original signal which provide more information than original signal. There are many ways of performing this task in literature like Fourier transform, STFT, FFT etc., but none of them are able to describe the time and frequency information of the waveform at a time which leads to development of a new transform to resolve above problem called as wavelet transform. The term 'wavelet' refers to an oscillatory vanishing wave with time-limited extend, which has the ability to describe the time-frequency plane, with atoms of different time supports. Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. They represent a suitable tool for the analysis of non-stationary or transient phenomena [30]. The STFT provides the identical resolution at all frequencies, but the use of wavelet transform provided a significance advantage to analyse the different frequencies of the signals with different resolutions by the method called multi-resolution technique The wavelet transform proofs its usability to describe the properties of waveform that changes with time by dividing the waveform into segment of scale. Thus, wavelet transform is very much suitable for analysing the EEG signals. There is two type of wavelet transform (I) continuous wavelet transform (I) discrete wavelet transform.

### 2.2.1 Continuous Wavelet Transform

Continuous wavelet transforms (CWT) [31] is defined by the equation:

$$W(m,n) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{|m|}}\psi^*\left(\frac{t-n}{m}\right)dt \qquad 2.1$$

Where b act to translate the function across the x(t), and the variable a act to vary the time scale of the probing function, $\psi$. While the probing functions $\psi$ could be any number of different function it always takes on an oscillatory form and thus the term "wavelet". If n=0 and m=1, then the wavelet is in natural form, which is termed as mother wavelet; that is, $\psi_{1,0}(t) \equiv \psi(t)$. A mother wavelet is shown in Figure (2.1) along with some of its family member produced by dilation and contraction.

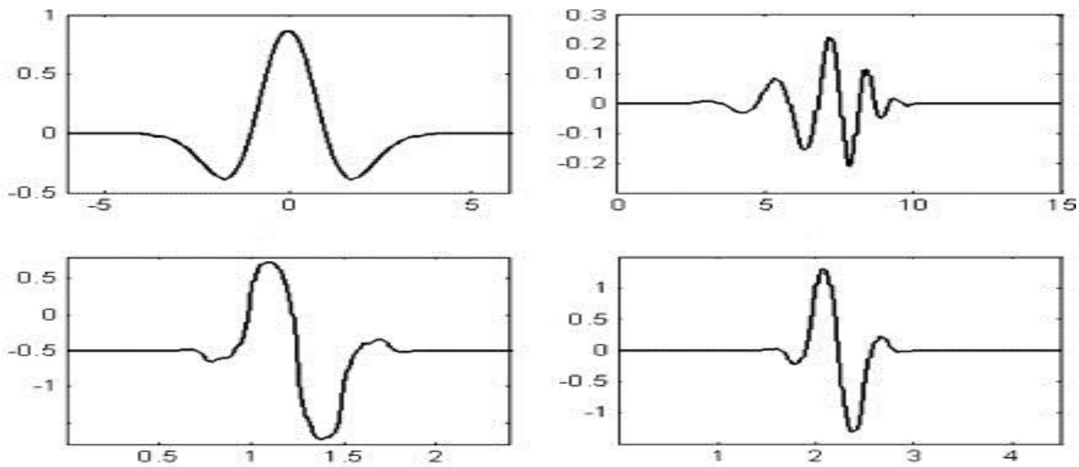


Figure 2.1: A mother wavelet (a=1) with two dilation (a=2 and a=4) and one contraction (a=0.5)

The wavelet shown is the popular morlet wavelet, and is defined by the equation:

$$\psi(t) = e^{-t^2} \cos(\pi\sqrt{\frac{2}{In2}}t) \quad 2.2$$

The wavelet coefficient W (m, n) describe the correlation between the waveform and wavelet at various translation and scale: similarly between the waveform and the wavelet at a given combination of scale and position, m, n. In the other way, the coefficient provide the amplitude of series of wavelet, over the range of scale and translation, that would need to be added together to reconstruct the original signal [31].

### 2.2.2 Discrete wavelet transforms (DWT)

There is a problem of redundancy in CWT, it provides oversampling of original signal and the coefficient generated in this process is more than it actually required to specify the signal. This redundancy create problem for recovery of original signal. To overcome this problem discrete wavelet transform is used which produce the minimum number of coefficient required to reconstruct the original signal. Discrete wavelet transform (DWT) provide multi-resolution analysis of a signal i.e. it gives information about signal in the time domain as well as frequency domain which is particularly useful in non-stationary signal like EEG. DWT of EEG signal is obtained by passing it to the series of filters. These filters are both low pass and high pass filters. Here the DWT analyses the EEG signal at various frequency bands by decomposing the signal. It employs the two sets of functions like scaling and wavelet functions, which are associated with the low pass and high pass filters respectively [32]. The original EEG signal is passing through the half band high pass and low pass filters. Then the signal sub sampled by factor 2, simply to discard other samples present in the signal [33]. This filtering approach can be implemented by following equation

$$D_{J+k}(i) = \sum_{l=0}^{L-k} g(k) A_j (2i + k) \quad 2.3$$

$$A_{J+k}(i) = \sum_{l=0}^{L-k} h(k) A_j(2i+k) \qquad 2.4$$

Where $j$ = 0,1,2,....., $M$ is the decomposition levels, initially which is equals to original sequence x(n). DWT has ability to extract the important information from EEG signals in the time-frequency domain which helps the characterization and detection process [34] [35].

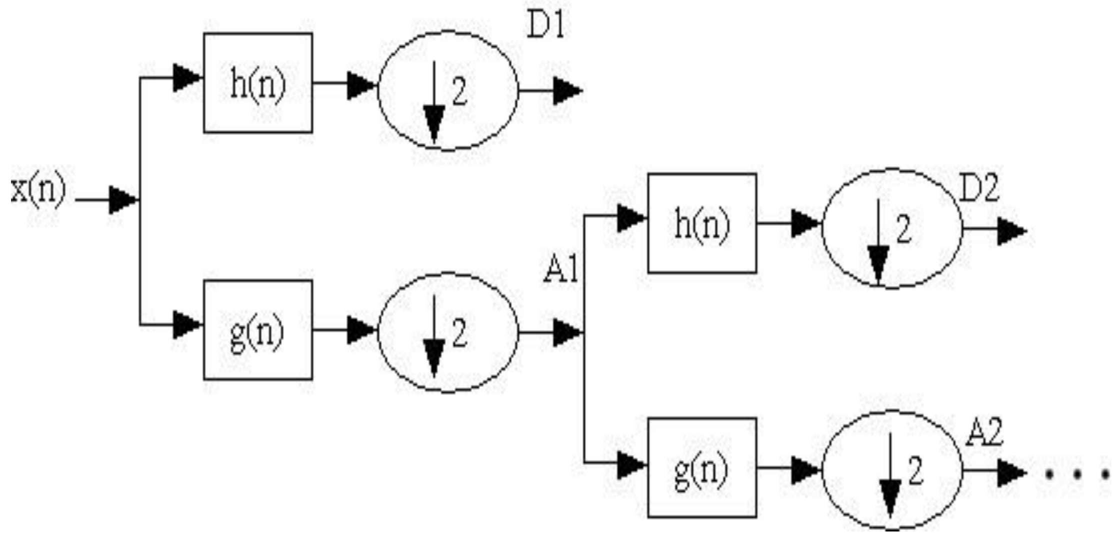


Figure 2.2: Sub band decomposition of discrete wavelet transforms implementation

The Discrete wavelet transform is performed by successive low pass and high pass filtering of the input signal x(n). The output of LPF and HPF filter is then 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. At each level, the outputs of the high pass filter after down sampled gives the detail coefficients Di and that of low pass filter gives the approximation coefficients Ai. The low resolution component is further decomposed and the procedure is continued up to two levels as shown in Figure. 2.2.

## 2.3 Feature extraction

The different stage of P300 detection is: pre-processing and normalization, feature extraction, feature reduction and classification. Figure 2.3 shows the flow diagram of whole process.

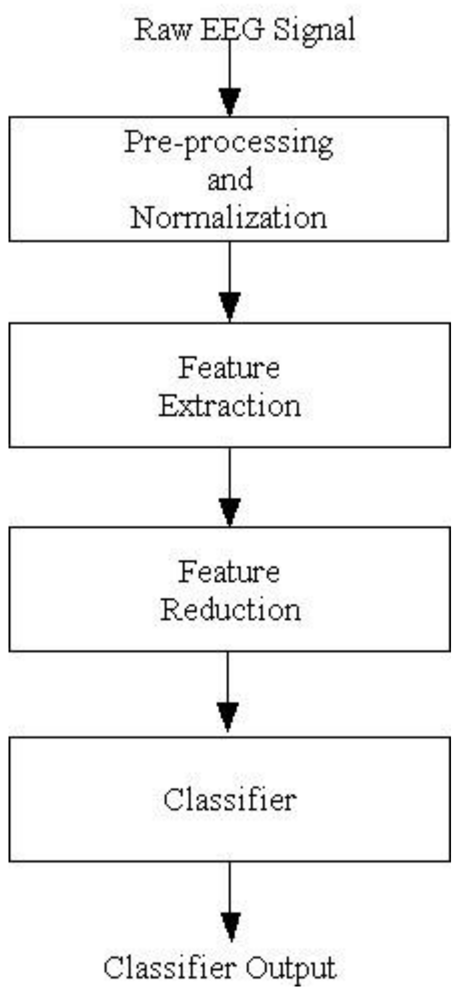


Figure 2.3: Flow diagram of P300 wave detection

### 2.3.1 Pre-Processing and Normalization

Since we only require the part of the recorded EEG signal that occurs after the intensification thus for each channel we extract all data samples between 0 to 667 ms posterior at the beginning of an intensification. As we know that P300 appears about 300 ms after the stimulus so we assume that the window we are considering is large enough to acquire all time domain feature for detection of P300 ERP [19]. Thus here number of samples extracted for each intensification is 160, which represent the 667 ms EEG signal having 240 Hz sampling rate. Next each recorded EEG signal is normalized to -1 and +1 value. This is done by dividing each sample with the maximum absolute value.

$$Normalize(k) = \frac{Signal(k)}{\|Signal(k)\|_{\max}} \tag{2.5}$$

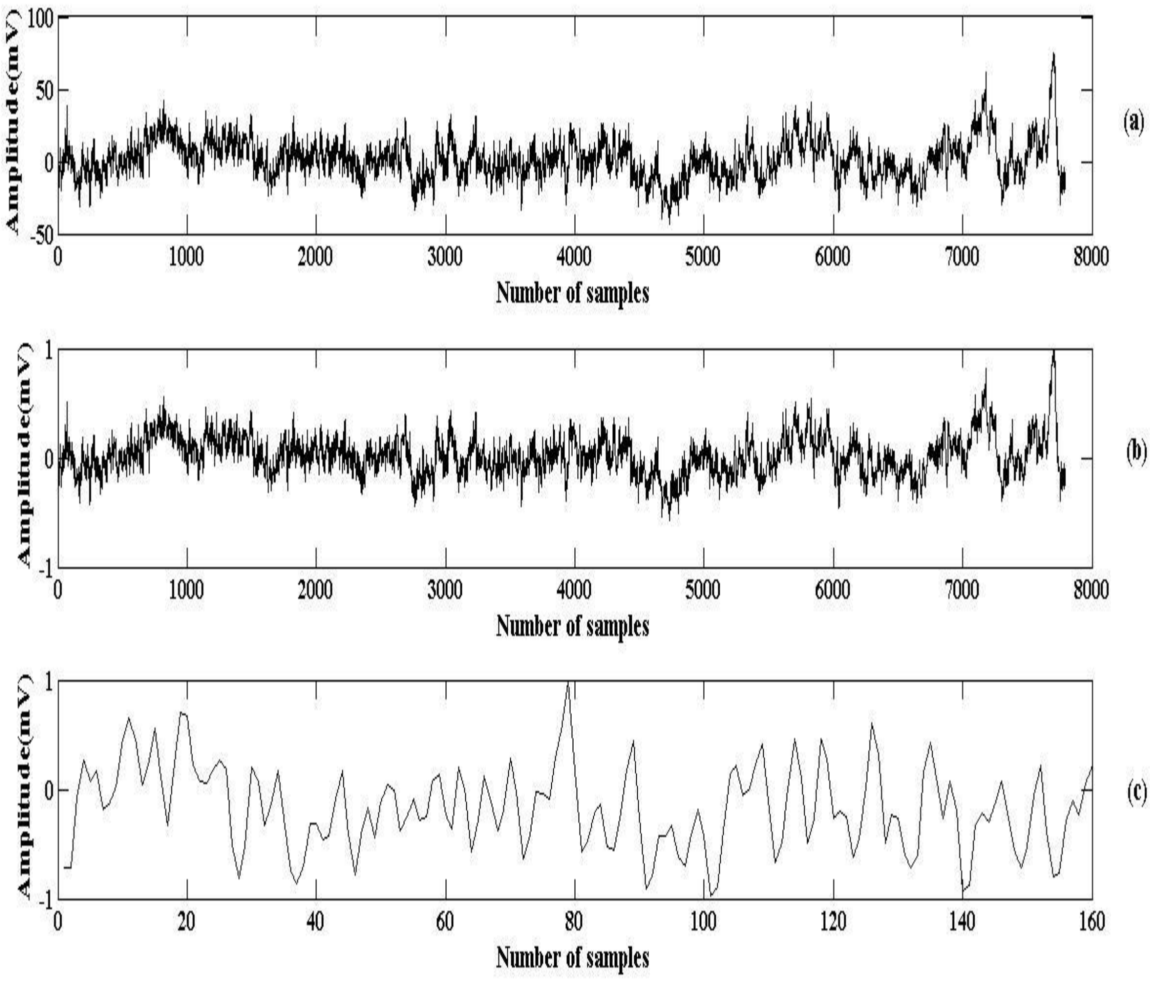


Figure 2.4: (a) Raw EEG signal (b) Normalised EEG signal (c) Obtained 667 ms ERP from raw EEG signal

### 2.3.2 Wavelet Features

In the present work, Daubechies wavelet [36] of order 4 and 2 level of decomposition is used for computation of wavelet coefficients. Each of the EEG signal is decomposed into two detail wavelet coefficients (D1 and D2) and two approximation wavelet coefficients (A1 and A2). Since P300 wave is most effective in the frequency range 0.1 to 20 Hz so here we decompose the EEG signal up to second level. Figure 3.5 and 3.6 show the waveform of number of wavelet coefficient at each level for two classes i.e. P300 and No-P300 respectively.

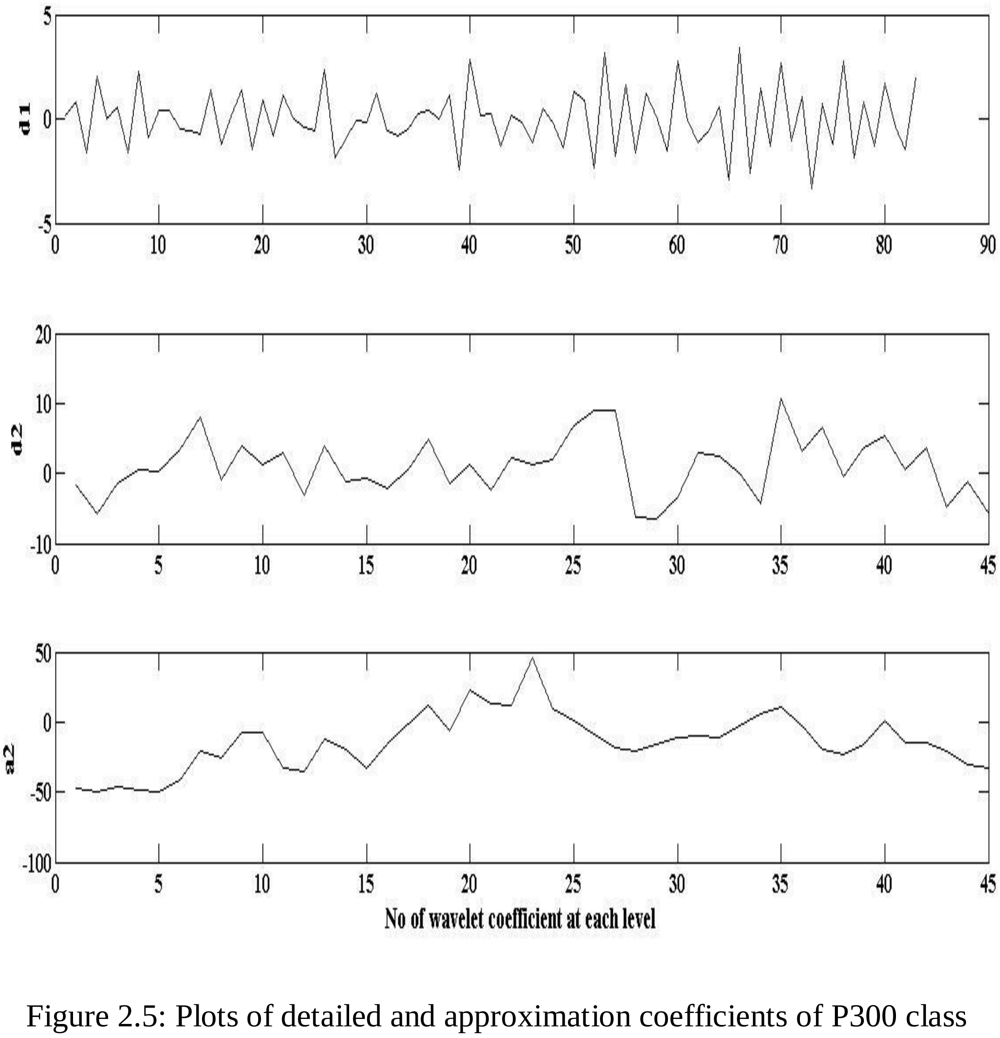


Figure 2.5: Plots of detailed and approximation coefficients of P300 class

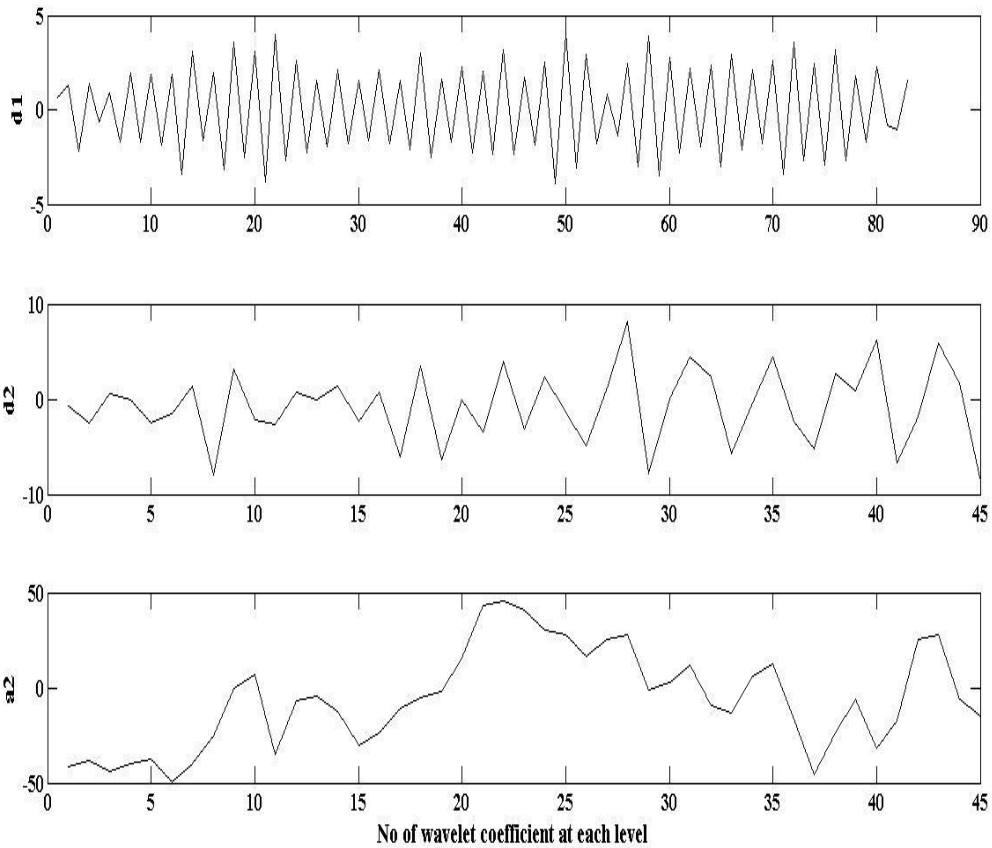


Figure 2.6: Plots of detailed and approximation coefficients of No-P300 class

Now *'wrcoef'* is used to reconstruct the coefficients of a one-dimensional signal which gives 160 wavelet coefficients for single visual stimuli for channel single. To reduce the size of the feature we divide the length of the approximate coefficient into 5 equal segments, after that we calculate two statistics parameter

(i) Maximum of wavelet coefficient in each sub band

(ii) Minimum of wavelet coefficient for each sub band.

Thus for each post stimulus signal we got 10 features for single channel and thus for 64 channels we have total of 640 = 64×10 feature vector.

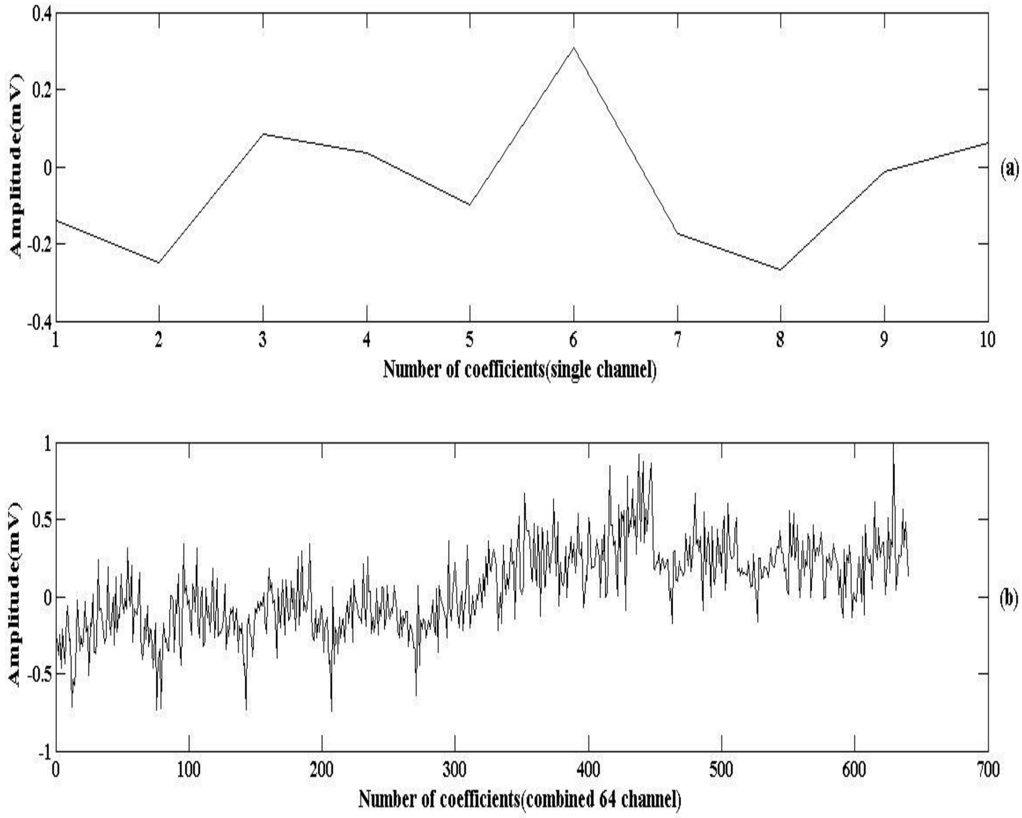


Figure 2.7: Wavelet feature (a) single channel (b) combined 64 channel

## 2.4 Feature Reduction Using PCA

A common problem in machine learning is feature selection. Feature selection means transforming the data space into feature space which has exactly the same dimension as the original data space, however transformation is designed in such a way that data set undergoes a dimensionality reduction. PCA is a well-established technique for feature extraction and dimensionality reduction. It is based on the assumption that most information about classes is contained in the directions along which the variations are the largest. The most common derivation of PCA is in terms of a standardized linear projection which maximizes the variance in the projected space. PCA is useful for data compression, by reducing the number of dimensions, without much loss of information [37].

We can perform PCA operation by following the steps mentioned below

Step1: Acquire the input data to be process

Step2: Obtain mean of the input data.

Step3: Subtract the mean from each of the input data dimension.

Step4: Calculate the covariance matrix. The required Covariance can be calculated from the equation:

$$\operatorname{cov}(x, y) = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1} \qquad 3.6$$

Where $\bar{X}$ is the mean of data set.

Step5: Next find the Eigen Vectors and Eigen Value of the Covariance matrix obtained in above step.

Step6: Now choose the components to form the feature vector and form the transformation matrix w.

Transformed data = w * Row Data

Where w is the matrix with eigenvectors in the columns transpose and Row data is the mean adjusted transpose. An assumption made for feature extraction and dimensionality reduction by PCA is that most information of the observation vectors is contained in the subspace spanned by the first m principal axes, where m < p for a p-dimensional data space. Therefore, each original data vector can be represented by its principal component vector with dimensionality m [37].

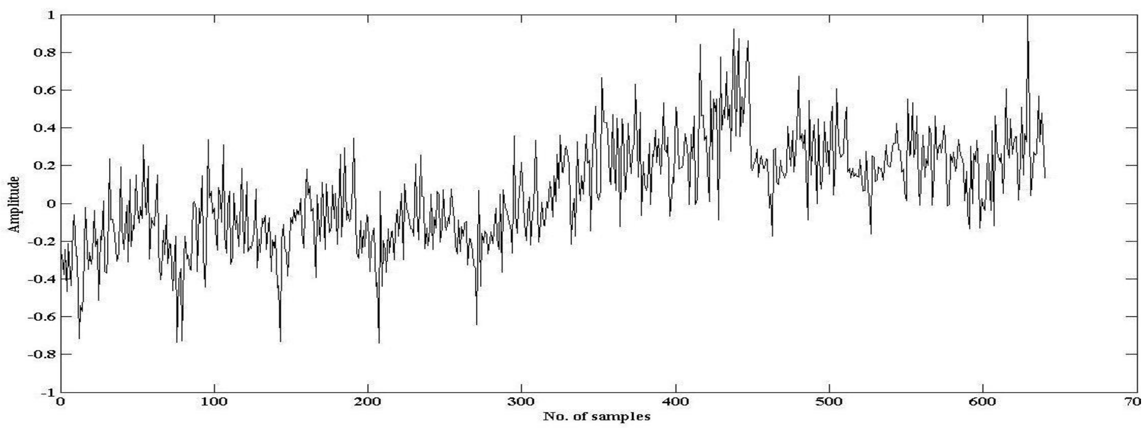


Fig. 2.8: Wavelet feature before PCA

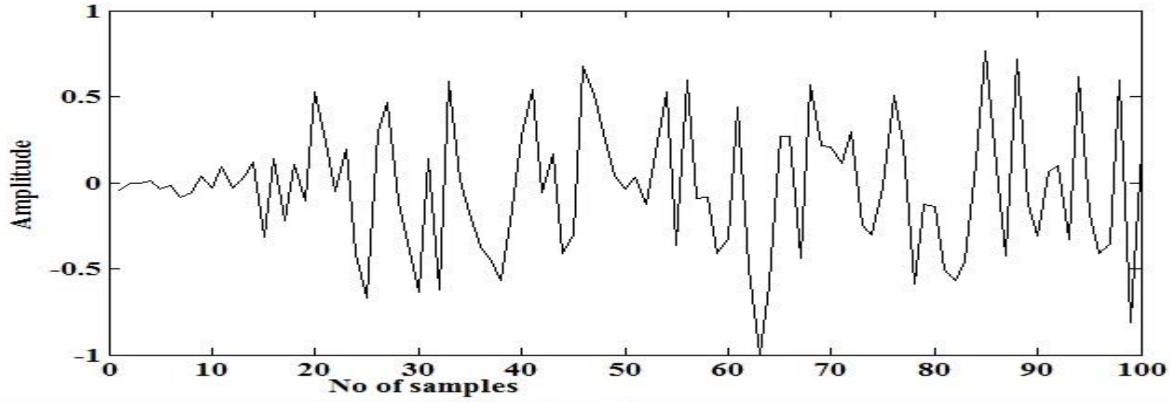


Fig 2.9: Reduced wavelet feature of P300 wave after PCA

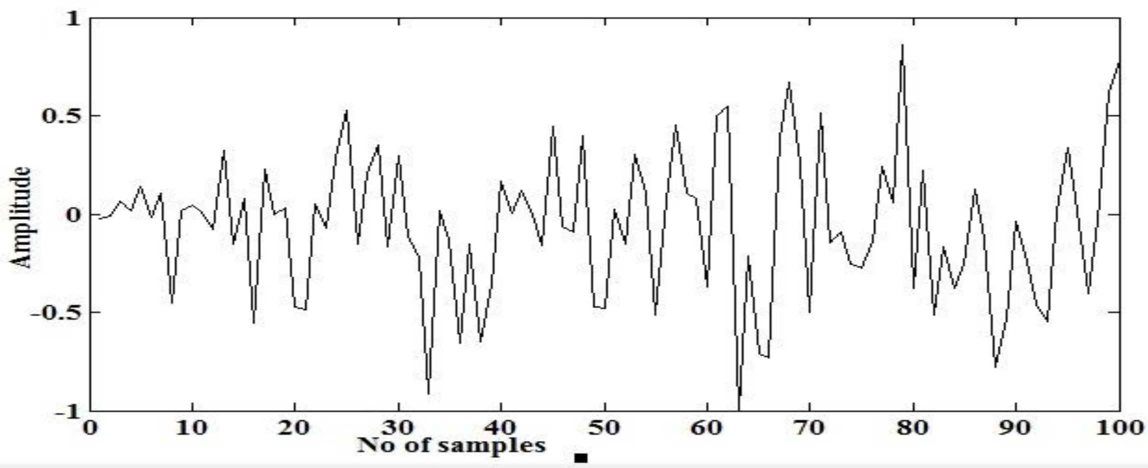


Figure 2.10: Reduced feature of No-P300 after PCA

In this case the feature vector obtained from discrete wavelet transform is of $1 \times 640$ dimension which is quite large input dimension for classifier. So here PCA is applied to reduce the feature dimension up to $1 \times 100$ for each stimulus response. Figure 2.9 and 2.10 shows the PCA reduced feature vector of target and non-target character.

# Chapter 3

# Classification

## 3.1 Introduction

The purpose of the BCI applications is to map the neurophysiologic signals to basic actions. This is performed with machine learning algorithms using any classification methods by creating a classification model. Depending on the classification problem the classification methods can be either parametric or non-parametric. However, the common approach in BCI problems is to train the learning algorithm first within a training phase in which the subject is asked, for example in Spelling Paradigm, to focus on a character that is known by the algorithm. This is a supervised learning methodology that constitutes a major topic in pattern recognition.

### 3.2 Classification Problem

The detection of the target character in P300 Speller requires the determination of the target row and column intensifications. There are 12 classes for a 6x6 speller matrix (6 row and 6 column intensification classes). There are two groups, namely the row and column groups since the target character should be at the intersection of only one row and one column. Hence, there are 6 classes in each separate classification group. Here, the problem may seem to be a 6 class classification task at first for each group. However, since the aim is to classify the stimulations as target (P300) or non-target (no-P300), the task becomes binary classification problem. The presence of one target element out of six elements in a trial is just a constraint or a priori information supplied by the problem.

## 3.3 Artificial Neural Network

Artificial neural networks are a mathematical model inspired by the functional aspect of biological structure of neurons commonly known as neural networks. ANN is particularly very much useful in the classification of EEG signal in literature because of its well known properties like; its structure is massively distributed in parallel compared to the other artificial intelligence method where information processing is sequential. Generalization property of the ANN makes it suitable for pattern recognition, generalization means its ability to learn from surrounding and experience and make it use in future to produce an adequate response to unknown stimulus which is related to acquire knowledge. Having this property produces some other property like adaptive learning, self-organization, fault-tolerance etc. This entire feature makes ANN able to give proper solution of complex problem normally difficult to solve by traditional approximation. The basic processing unit of brain is neuron which works identically in ANN. The fundamental unit of neural network is neurons which are interconnected with each other via 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 [38]. A neuron is an information-processing unit that is fundamental to the operation of a neural network.

### 3.3.1 Models of Neuron

A neuron is a fundamental information processing unit of neural network. Figure 3.1 shows the model of neuron. The basic element shows in the model are;

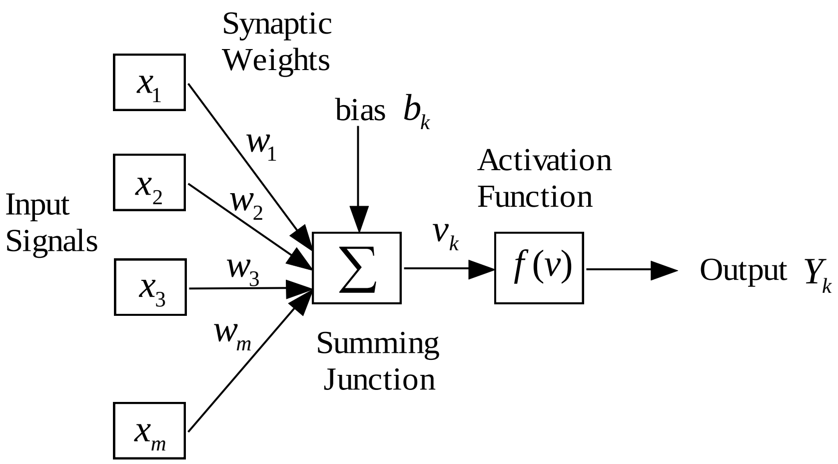


Figure 3.1: Nonlinear model of a neuron

(i) A set of connection link or synapses characterised by its weight. A signal $x_j$ at input of synapse j connected to neuron k multiplied by weight $w_{kj}$.

(ii) An adder to summing the input signal multiplied by its synaptic weight which is a linear combination.

(iii) An activation function for putting limit in the output of neuron. Typically, the normalized amplitude range of the output of a neuron is written as the [0, 1] or alternatively [-1, 1].

(iv) Bias denoted by $b_k$, which has effect of increasing or lowering the net input of the activation function

In mathematical term we can write the neuron *k* by following equations;

$$v_k = \sum_{j=0}^{m} w_{kj} x_j \qquad 3.1$$

and

$$y_k = \varphi(v_k) \qquad 3.2$$

Or in matrix form

$$v_k = \lfloor w_{k0}\ w_{k1} \ldots w_{kp} \rfloor \begin{bmatrix} x_0 \\ x_1 \\ \cdot \\ \cdot \\ x_p \end{bmatrix} = w_k^T x \qquad 3.3$$

### 3.3.2 Network Architecture

The way in which the fundamental unit i.e. neurons is structured is directly linked to to the learning algorithum used to train the network. In gereral there are three types of ANN architecture: (i) Single layer feedforward network (ii) Multilayer feedforward network (iii) Recurrent network. For the classification problem we are dealing here uses multilayer feedforward network out of three available architecture.

- **Multilayer Feed-forward Network**

A multi-layer feed-forward network is widely used for classification of EEG signals. It has two or more layers of units, with the output from one layer serving as input to the next. There are no connections within a layer. The input layer has neurons which are merely ’fan out’ units, where N equals the number of classification inputs, no processing takes place in these units. The layers with no external output connections are referred to as hidden layers, while there are M neurons in the output layer, where M equals the number of classification outputs. In most cases, a feed-forward NN with one hidden layer of units is used with a sigmoid activation function of the units [39] [40].

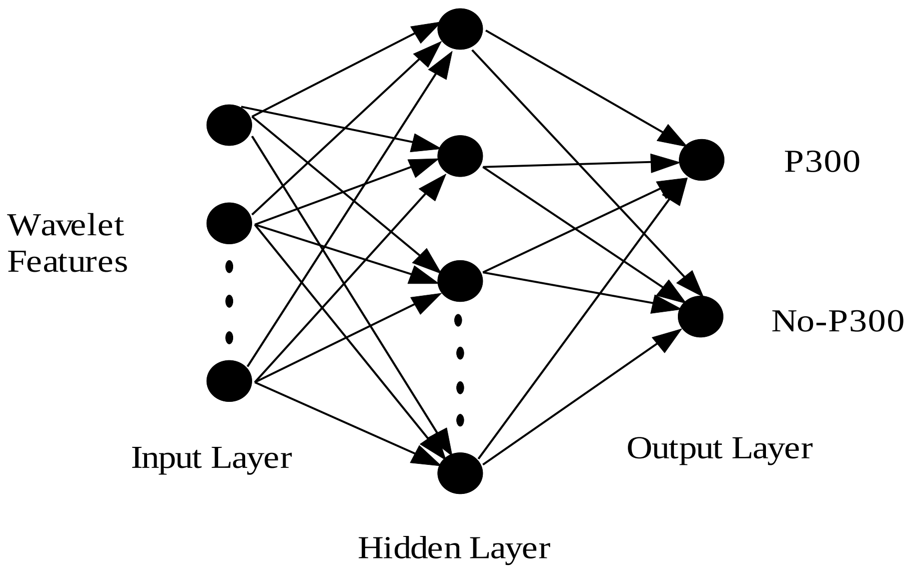


Figure 3.2: Multilayer feedforward network

As presented in Figure 3.2 the input neurons in the first layer, distribute the inputs to neurons in subsequent layers. In the next layers, each neuron sums its inputs and adds a bias or threshold term to the sum and non-linearly transforms the sum to produce an output

### 3.3.3 Multilayer perceptron

One of the most important models in the ANN is multilayer perceptron which is the simplest form of a neural network and finds very useful in classification of linearly separable patterns. It has 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 hyperplane between the classes. The proof of convergence of the algorithm is known as the *perceptron convergence theorem.* Multilayer perceptron 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 back propagation training with generalized delta learning rule is an iterative gradient algorithm designed to minimize the root mean square error between the actual output of a multilayered feed-forward NN and a desired output. Each layer is fully connected to the previous layer, and has no other connection [37].

Back propagation algorithm;

1. First initialise the weight and biases to some real random number.
2. Specify input vector $x(1), x(2),...., x(N)$ and corresponding output vector $d(1), d(2),....., d(N)$, one pair at a time, where N is the number of training pattern.
3. Calculate the actual output $y_1, y_2,....., y_{NM}$ by using the equation

$$\left(\sum_{j=1}^{N_{M-1}} w_{ij}^{(M-1)} x_j^{M-1} + b_i^{M-1}\right), \ i = 1,...., N_{M-1} \qquad 3.4$$

4. Adaption of weight $(w_{ij})$ and biases $(b_i)$

$$\Delta w_{ij}^{l-1}(n) = \mu . x_j(n) . \delta_i^{l-1}(n)$$

$$\Delta b_i^{l-1}(n) = \mu . \delta_i^{l-1}(n)$$

Where

$$\delta_i^{l-1}(n) = \begin{cases} \varphi^t\left(net_i^{l-1}\right)\left[d_i - y_i(n)\right], \; l = M \\ \varphi^t\left(net_i^{l-1}\right)\sum_k w_{ki} . \delta_k^{(l)}(n), \; 1 \le l \le M \end{cases}$$

In which $x_j(n)$ = output of node j at iteration n, l represent layer, k is the number of output node, M is output layer and $\varphi$ is activation function where as $\mu$ represent the learning rate. To achieve faster convergence with minimum oscillation, a momentum term is added to the basic weight updating equation. After completing the training procedure of the neural network, the weights of MLP are frozen and ready for use in the testing mode [37].

# 3.4 Support Vector Machine Classification

Support vector machine (SVM) is a powerful algorithmic for classification and regression [41] [42]. A classification means separating the datasets into training and testing sets. Each sample in the training set contains one “target value” and several “attributes". The goal of SVM classifiers generates a model or a decision surface based on the training data, which recognizes the target value of the testing dataset.

## 3.4.1 Optimal Hyperplane

Let’s consider the two class classification case described in section 3.2, where the set of stimulus response are to be separated by a discriminating function. For such kind of problem, one can find infinitely many separating hyperplanes hence many functions shown in Fig 3.3

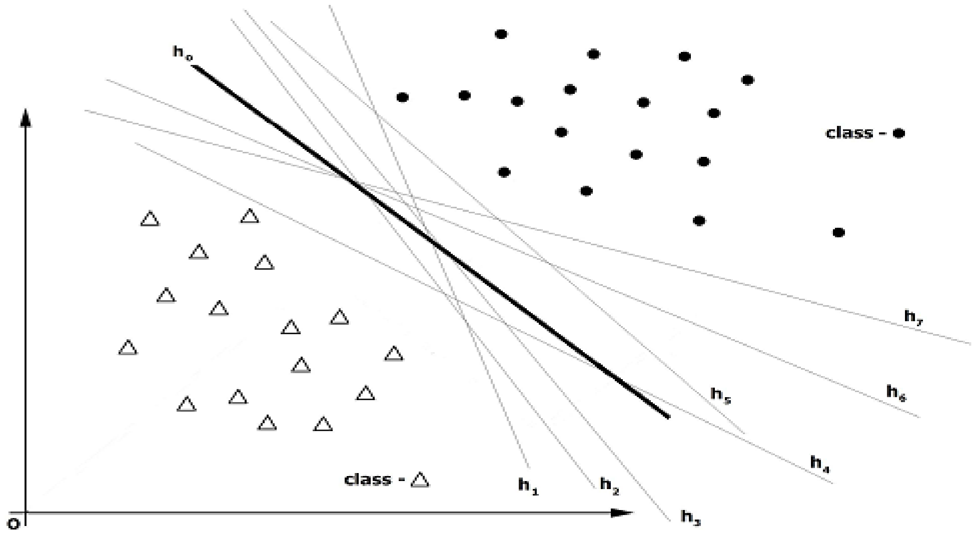


Figure 3.3: Several solutions of the discriminating function for a linearly separable 2 category classification problem.

However, among all these hyperplanes, there is only one which provides the maximum margin of separation between the classes. That is, SVM tries to find the hyperplane which maximizes the distance between the hyperplane and the nearest samples in each class [42]. The optimum separating hyperplane is illustrated in Figure 3.4.

The nearest element to the hyperplane lying on $H_1$ and $H_2$ in Figure 3.4 are called the support vectors which give the name to the method [43]. The solution procedure of the optimum separating hyperplane (OSH) reveals these elements and once these are found the classification in the test phase is done by using the support vectors [6].

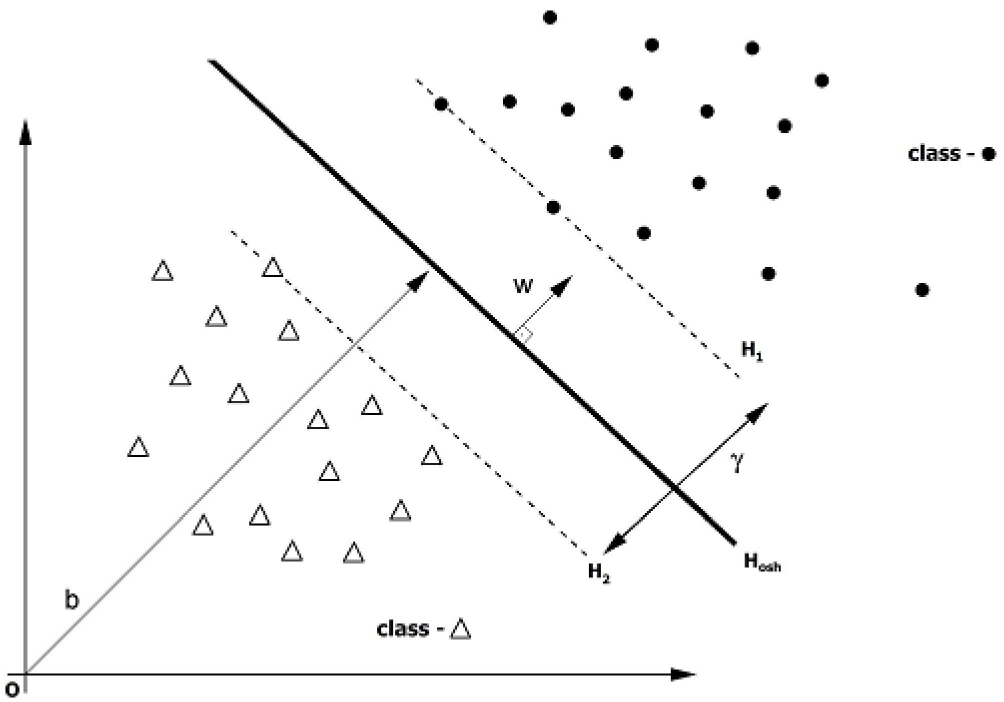


Figure 3.4: Optimum Separating Hyperplane (OSH) of SVM for a two class case. The closest samples to OSH from each class are called the support vectors [6].

### 3.4.2 Architecture of SVM

The Architecture of SVM is shown in figure 4.3 where $x_1, x_2, x_3......x_i$ the input vector of size m. $K(x,x_1), K(x,x_2)......., K(x,x_i)$ $K(x,x_1), K(x,x_2)......., K(x,x_i)$ are the hidden layer of the feed-forward neural network, w is the weight vector having weighted elements $w_1, w_2, w_3......., w_m$, b is the bias defined in the n dimensional space. Y is the output which makes decision for classification.

The decision boundary for the classification purpose is defined as

$$w^T \Phi(x) + b \qquad 3.5$$

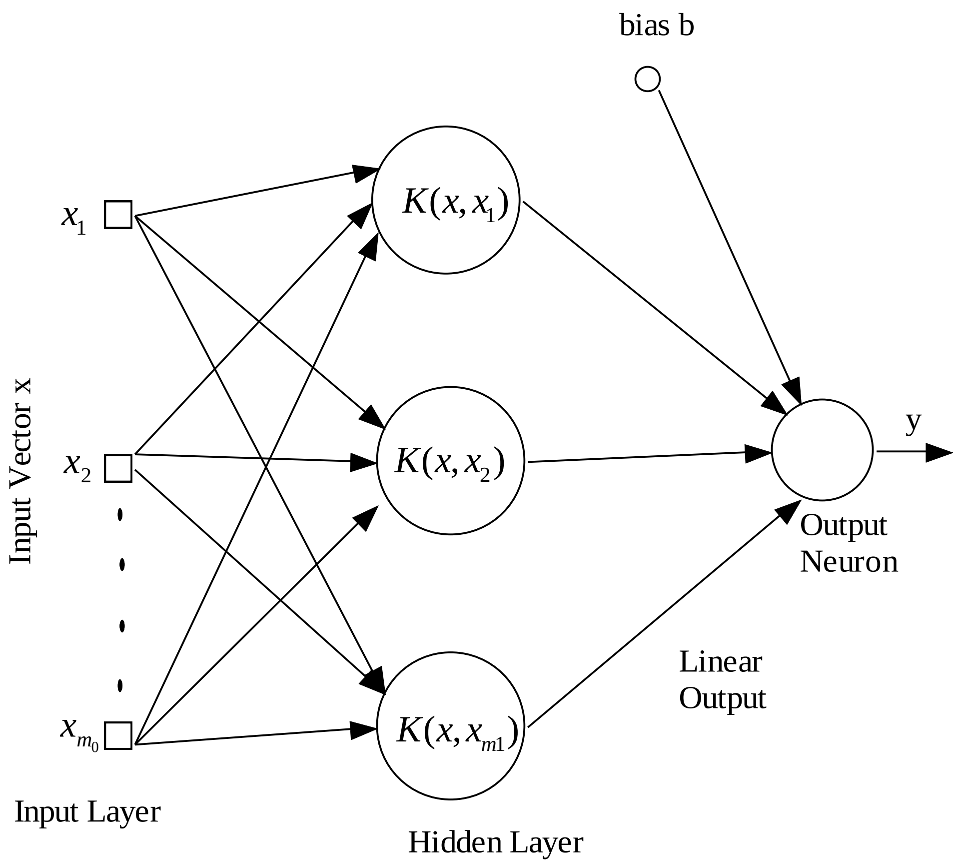


Figure 3.5: Architecture of SVM

Where $w \in R^n$ is the weight matrix, $\Phi(x)$ is the kernel function of x and b is the position of hyper plane in real dimension space R, $b \in R$ .

$$w^T \Phi(x_i) + b \geq +1, \quad if\ y_i = +1$$

$$w^T \Phi(x_i) + b \leq -1, \quad if\ y_i = -1$$

For training set of instance label pairs $\left(X^i, y^i\right)$ where i=1 to number of inputs m, where $X^i \in R^n$ , $y^n \in \{-1,1\}$ both defined in real dimension space.

Which is equivalent to $y_i\left[w^T \Phi\left(xi\right) + b\right] \geq 1$, i=1… m

The optimal hyper plane should satisfy following criteria

$$y_i\left[w^T\Phi\left(x_i\right)+b\right]\geq 1 \qquad 3.6$$

$$0\leq y_i\left[w^T\Phi\left(x_i\right)+b\right]<1 \qquad 3.7$$

$$y_i\left[w^T\Phi\left(x_i\right)+b\right]<0 \qquad 3.8$$

The equation (3.6) states that the vectors that fall outside the band and are correctly classified. The equation (3.7) states that the vectors falling inside the band and is correctly classified and equation (3.8) states that the vectors are misclassified vectors. All the above three cases can be restricted by introducing a new set of non negative slack variable $\xi_i$

$$y^i\left[w^T\Phi\left(x_i\right)+b\right]\geq 1-\xi_i$$

And $\xi_i \geq 0$

### 3.4.3 SVM Formulation

The equation of the decision surface in the form of hyperplane is

$$w^t x+b=0 \qquad 3.9$$

Where x is the input vector, w in an adjustable weight and b is bias. To obtain OSH it is required to maximize the margin $\gamma$. To begin let's say one set of class labels as y= {-1, 1}, then $y_i\left(w^t x_i+b\right)\geq 1$ becomes a constraint of this optimization problem. This constraint comes from the restriction that

$$\min\left|\left(w^t x_i+b\right)\right|=1 \qquad 3.10$$

which simplifies the formulations in the problem. Thus margin γ is related to the weight vector w with:

$$\gamma = \frac{2}{\|w\|} \qquad 3.11$$

Equation 4.11 states that maximizing the margin of separation between binary classes is equivalent to minimizing the Euclidean norm of the weight vector w.

So the quadratic support vector optimization problem is formulated as follows:

$$\text{Minimize } \frac{1}{2}\|w^2\|$$

$$\text{s.t. } \; y_i\left(w^t x_i + b\right) \geq 1 \qquad i = 1,2,\ldots\ldots, N \qquad 3.12$$

Where N is the number of element in the training set.

Above equation (3.12) can provide OMS solution only for linearly separable classes. However in real problems the data is inseparable even with the feature extraction methods and higher dimension mappings. To overcome this problem, we introduce a penalty function and positive slack variables to modify the equation (3.12) and to get required optimization problem. [42] [44].

$$\text{Minimize } \frac{1}{2}\|w^2\| + c\sum_{i}^{N}\xi_i$$

$$\text{s.t. } y_i\left(w^t x_i + b\right) \geq 1 - \xi_i \qquad i = 1,2,\ldots., N \qquad 3.13$$

$$\forall i$$

where $\xi_i$ 's are the slack variables and C is the regularization parameter. This modification provides a more flexible situation which is called as the “soft-margin” SVM optimization problem [45]. To simplify the solution of OSH, the cost function in (3.5) is expressed in the Lagrangian dual form [46]

$$\text{maximize } L_D \equiv \sum_i a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i^t x_j$$

$$\text{s.t. } 0 \le a_i \le C$$

$$\sum_i a_i y_i = 0$$

From above equation we can get the solution in term of Lagrange multiplier $a_i$ for all training point $x_i$ and the non-zero solution for $a_i$'s represent the coefficient of the support vectors.

Thus, the solution for the weight vector w of the OSH is :

$$w = \sum_i^{N_s} a_i y_i x_i \qquad 3.15$$

where $x_i$'s and $N_s$ represent the support vectors and the number of the support vectors respectively.

The solution for the threshold or the bias term b can be found from the Karush-Kuhn-Tucker (KKT) conditions for the primal Lagrangian form which states

$$a_i\left[y_i\left(x_i^t w + b\right) - 1\right] = 0, \qquad \forall i \qquad 3.16$$

The slack variable $\xi_i$ is removed here by using the fact that $0 \le a_i \le C$ this constraint implies that $\xi_i = 0$ when the Lagrange multipliers are in this range. Therefore, from (3.16) one can use all the training points to compute the bias term b for which $0 \le a_i \le C$ is satisfied. Since there are at most N equations in (4.8) one may find N different solutions for b and the common approach is to average all these solutions to compute the bias vector [42].

### 3.4.4 Kernel Methods

Normally the Kernel operator used in SVM is the dot product of the two vectors. However, other transformations can also be used to generalize the SVM for the nonlinear case. The basic idea is to map the vectors into a higher dimensional space which provides some advantages in cases where the data is linearly inseparable. The mapping is not subjected to any constraints, so kernel method can conduct to infinite dimensional space for the purpose of classification. A tool called kernel trick is used in order to map the patterns in the higher dimensional space. The selection of appropriate kernel depends upon type of patterns available. There are following choices for kernel function:

(i) Polynomial Kernel:

$$K(x_i, x_j) = \langle x_i, x_j \rangle^d \tag{3.17}$$

(ii) Gaussian radial basis function

$$K(x_i, x_j) = e^{\frac{\|x_i - x_j\|^2}{2\sigma^2}} \tag{3.18}$$

(iii) Exponential Radial basis function

$$K(x_i, x_j) = e^{\frac{\|x_i - x_j\|}{2\sigma^2}} \tag{3.19}$$

(iv) Dirichlet Kernel:

$$K(x_i, x_j) = \frac{\sin\left(\left(N + \frac{1}{2}\right)(x_i - x_j)\right)}{2\sin\left(\frac{x_i - x_j}{2}\right)} \tag{3.20}$$

(v) Sigmoid Function:

$$K(x_i, x_j) = \tanh(kx_i.x_j - \delta) \tag{3.21}$$

# Chapter 4

# Results and Conclusion

## 4.1 Experimental Result of ANN and SVM Classifier

As explained in section 1.6, Data set II of the third BCI competition is used here to evaluate the proposed methodology. In this work the main goal is to detect the P300 wave accurately. So P300 detection is a binary class classification problem, one class refers to P300 stimuli and second class refers to no P300 stimuli. To solve this classification problem here we are using two very efficient classification method namely artificial neural network (ANN) and support vector machine (SVM). For detection of P300 wave we are using data-set named 'Subject A Train' and 'Subject B Train' from the Data set II of third BCI competition available online publically at BCI competition web page [27]. Each subject composed of 85 characters and whose target character is known. Each character epoch is supposed to contain two P300 signals, one for row flash and one for the column flash. So in the total database, the number of P300 is 85×2×15 i.e.2550 and no P300 is 85×10×15 i.e. 12250 for each subject.

In classifier training and testing, to compensate the unbalance between target (P300) and non-target (no P300) samples (the number of non-target samples is five times to the number of target samples), a randomized 1/5 selection was performed on the non-target feature dataset. Thus for each of the two subjects we have total 2550 number of P300 stimulus and 2550 numbers of no P300 stimulus. In this evaluation we are using 50% data for training purpose and remaining 50% data for testing purpose.

As explained above here the data base size for each subject is having 2550 number of P300 and 2550 number of No-P300 response and each one of these response has feature dimension of 1×100 which is off course obtained after applying the feature reduction technique called PCA to the feature dimension obtained after wavelet transform feature extraction to the original signal which is 1×640. Since here 50% of database is used for training purpose and rest of 50% is used for testing purpose thus we have 1275 number of P300 and 1275 number of No-P300 for training phase and testing phase. Table 5.1 shows the classification result obtained from ANN and SVM classifier. It is clear from the result obtained that for both the subject A and B, the SVM classifier proofs its superiority over ANN classifier

Table 4.1: Confusion Matrix for ANN and SVM classifier for Subject A and B

| Classifier | Subject | Desired Result | Output | |
|---|---|---|---|---|
| | | | P300 | No-P300 |
| ANN | A | P300 | 1029 | 314 |
| | | No-P300 | 246 | 961 |
| SVM | A | P300 | 1271 | 228 |
| | | No-P300 | 4 | 1047 |
| ANN | B | P300 | 1089 | 242 |
| | | No-P300 | 186 | 1033 |
| SVM | B | P300 | 1271 | 196 |
| | | No-P300 | 4 | 1079 |

Performance of the classifier is evaluated on various statistical parameters like

*Precision:* The percentage of positive predictions (P300) that is correct and is given by $TP/TP+FP$

*Sensitivity:* The percentage of positive labelled instances (P300) that were predicted as positive (P300) and are calculated as $TP/TP+FN$

*Specificity:* The percentage of negative labelled instances (No-P300) that were predicted as negative (No-P300) and are given by $TN/TN+FP$

*Accuracy:* The percentage of predictions (P300 and No-P300) that is correct and is given by $(TP+TN)/(TP+TN+FN+FP)$

Table 4.2: Evaluation parameter of different classifiers

| Subject | Classifier | Sensitivity (%) | Accuracy (%) | Specificity (%) | Precision (%) |
|---|---|---|---|---|---|
| A | ANN | 76.61 | 78.03 | 79.61 | 80.70 |
| | SVM | 84.78 | 90.90 | 99.61 | 99.68 |
| B | ANN | 81.81 | 83.21 | 84.74 | 85.41 |
| | SVM | 86.63 | 92.15 | 99.63 | 99.68 |

The main observation from the performance criteria shown in table 5.2 is the large difference between the two subjects. For the same classifier the accuracy and sensitivity of subject B is much higher than subject A. The difference between the precision and sensitivity in relation to the methods can find a link between one of these measurements and can be used to next step of character prediction.

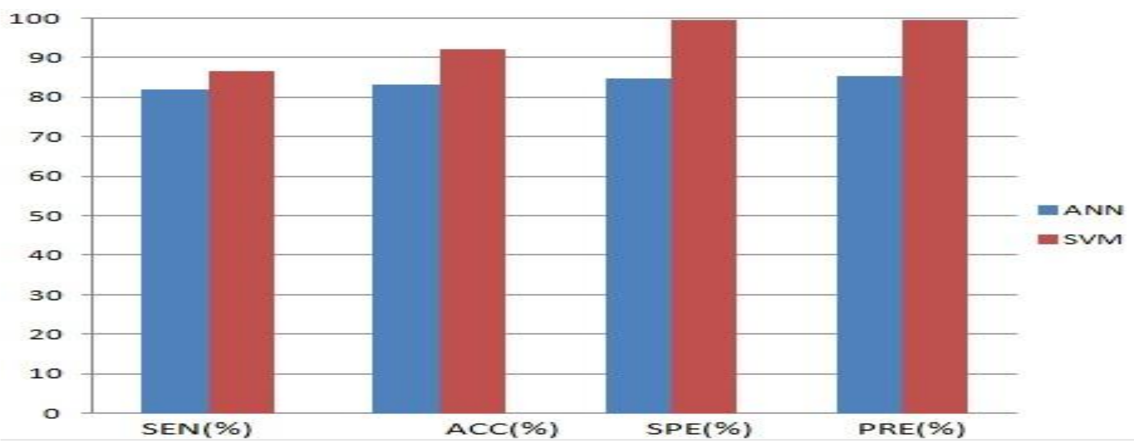


Figure 4.1: Performance comparison between ANN and SVM for subject A

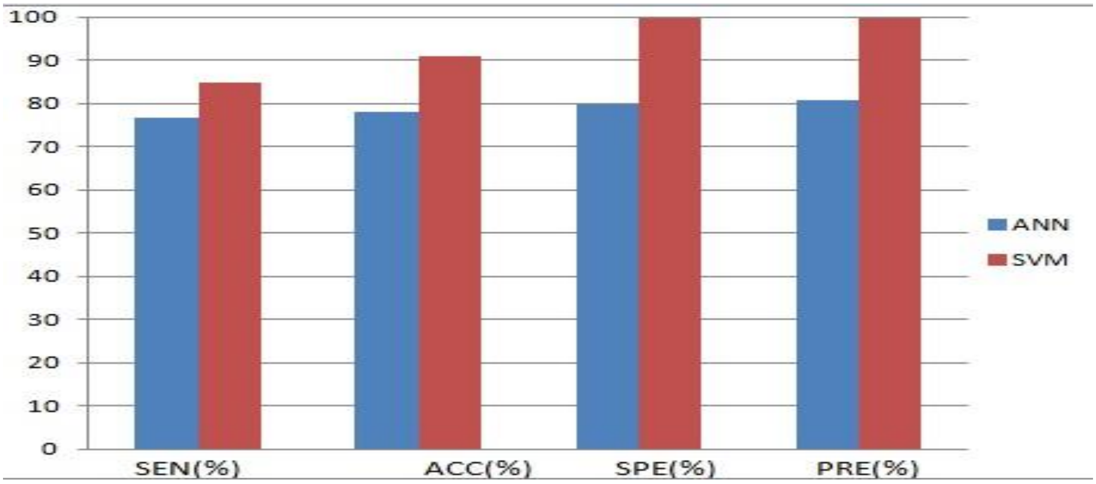


Figure 4.2: Performance comparison between ANN and SVM for subject B

Figure 4.1 and 4.2 shows graphical comparision analysis of various performance parameter for two subject based on ANN and SVM classifier. Off-course SVM is better in each parameter but the difference between percentage accuracy for subject B is more than subject B and similarily for other parameters, which shows that performance of detection of P300 also depend on the subject i.e. it varies user to user.

## 4.2 Conclusion

To bring the BCI system from the laboratory to the real world application it is required that its preformance should be as accurate as possible and the BCI system based on P300 speller untill and unless we are not able to detect the P300 correctly we can not convert the brain signal to actual command and thus the main goal of this thesis is the detection of P300 wave accurately. For this purpose discrete wavelet transform proofs itself a very robust feature extraction techniqe to extract the most releted feature from the original EEG signal obtained from brain. Since lose of any data recorded from brain is not tolerable so all the 64 channel data has been used for analysis purpose which certainly produce curse of dimensionality and makes classification process complex. To cope up this problem principal component analysis employed to reduce the size of feature which is to be used as input to the classifier and this is the one more step towards efficient detection of P300. The classifier result shows that SVM detect P300 more accurately then ANN this is because SVM has some advantage over ANN ae described in litrature like, SVMs have been developed in the reverse order to the development of neural networks (NNs). SVMs evolved from the sound theory to the implementation and experiments, while the NNs followed more heuristic path, from applications and extensive experimentation to the theory. Also In comparison with traditional multilayer perceptron neural networks that suffer from the existence of multiple local minima solutions, convexity is an important and interesting property of nonlinear SVM classifiers. Because of the above reason using SVM is better option for P300 detection.

## 4.3 Future Work

There is large scope of future work in this research, some of them are:

(i) As we have seen the accracy of P300 detection varies with the subject thus the result is somewhat limited to very less number of subject, to make it generlised result more set of testing is required and also inter subject variability can also be a fact to be consider.

(ii) After accurately detection of the P300, the result obtained can be utilised to predict the character that user focusing and convert the brain signal into actual command.

# Bibliography


[1] Bernhard Graimann, Brendan Allison, and Gert Pfurtscheller, “Brain–computer interfaces: A gentle introduction,” *The Frontiers Collection Springer Berlin Heidelberg*, pp.1-27, 2010.

[2] J.R Wolpaw, D.J. McFarland., T.M. Vaughan, G. Schalk, “The Wadsworth Center Brain-Computer Interface (BCI) Research and Development Program,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 1-4, June 2003.

[3] S.G. Mason, G.E Birch, “A General Framework for Brain-Computer Interface Design”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 1, pp. 70-85, Mar 2003.

[4] F. Cincotti, L. Bianchi, G. Birch, C. Guger, J. Mellinger, R. Scherer, R.N. Schmidt, O.Y. Suarez, G. Schalk, “BCI Meeting 2005 – Workshop on Technology: Hardware and Software,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no.2, pp. 128-31, June 2006.

[5] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, B. Arnaldi, “A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces,” *Journal of Neural Engineering*, vol. 4, pp. 1-13, Jan 2007.

[6] Hasan Balkar Erdoğan “A Design And Implementation Of P300 Based Brain-Computer Interface” Ph.D. thesis, The Graduate School Of Natural And Applied Sciences of Middle East Technical University, 2009.

[7] Hubert Cecotti, Axel Graser, “Convolutional neural networks for P300 detection with application to brain-computer interfaces,” *IEEE transactions on communications*, vol. 33, no. 3, pp. 433-445, Mar 2011.

[8] Saeid Sanei, and A. Jonathon Chambers. *EEG signal processing*. Wiley-Interscience, 2008.

[9] Michal Teplan, “Fundamentals of EEG measurement,” *Measurement science review*, vol. 2, no. 2, pp 1-11, 2002.

[10] Towle, L. Vernon, Bolanose Jose, Suarez Diane, Tan Kim, Robert Grzeszczuk "The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy." *Electroencephalography and clinical neurophysiology*, vol. 86, no. 1, pp. 1-6, Jan 1993.

[11] Nicolas-Alonso, Luis Fernando, and Jaime Gomez-Gil, “Brain computer interfaces, a review”, *Sensors*, vol. 12, no. 2, pp. 1211-1279, Jan 2012,

[12] Dean J Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, J. R. Wolpaw, “Toward enhanced P300 speller performance,” *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15–21, Jan 2008

[13] L. Maggi, S. Parini, L. Piccini, G Panfili, G. Andreoni, “A four command BCI system based on the SSVEP protocol” *IEEE International Conference on Engineering in Medicine and Biology Society,* pp. 1264-1267, 2006.

[14] Ben H. Jansen, A. Allam, P. Kota, K. Lachance, “An exploratory study of factors affecting single trial P300 detection,” *IEEE Transactions on Biomedical Engineering,* vol. 51, no. 6, pp. 975-978, June 2004.

[15] Min Ki Kim, and Sung-Phil, “Detection of P300 components using the Wiener filter for BCI-based spellers,”*8th Asian IEEE conference on Control (ASCC*), pp. 892-896, sep. 15-18, 2011.

[16] Gerardo Rosas-Cholula,“On signal P-300 detection for BCI applications based on wavelet analysis and ICA preprocessing,” *IEEE conference on Electronics, Robotics and Automotive Mechanics Conference (CERMA*), pp.360-365, 2010.

[17] Matthias Kaper, M. Meinicke, U. Grossekathoefer, “BCI competition 2003-data set IIb: Support vector machines for the P300 speller paradigm,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1073-1076, June 2004.

[18] Vladimir Bostanov, “BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram,” *IEEE Transactions on Biomedical Engineering,* vol. 51, no.6, pp. 1057-1061, June 2004.

[19] Alain Rakotomamonjy and Vincent Guigue, “BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller,” *IEEE Transactions on Biomedical Engineering,* vol. 55, no. 3, pp. 1147-1154, Mar 2008.

[20] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, H. Ritter, “BCI Competition 2003 – Dataset IIb: Support Vector Machines for the P300 Speller Paradigm,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1073 – 1076, June 2004.

[21] B. Blankertz, K.R. Müller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, N. Birbaumer, “The BCI Competition 2003: Progress and Perspectives in Detection and Discrimination of EEG Single Trials,” *IEEE Transactions on Biomedical Engineering*, vol.51, no.6, pp. 1044 – 1051, June 2004.

[22] Z.L. Lin, C.S. Zhang, “Enhancing Classification by Perceptual Characteristics for the P300 Speller Paradigm,” *Proceedings of the 2nd International IEEE EMBS Conference on Neural Engineering – Arlington/Virginia*, March 16-19, 2005.

[23] M. Thulasidas, C. Guan, J. Wu, “Robust Classification of EEG Signals for Brain-Computer Interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1, pp. 24-29, Mar 2006.

[24] D. Garrett, D.A. Peterson, C.W. Anderson, M.H. Thaut, “Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141-144, June 2003.

[25] B. Blankertz, G. Curio, K.R. Müller, "Classifying Single Trial EEG: Towards Brain Computer Interfacing," *Advanced. Neural Information Processing Systems*, vol. 14, pp. 157-64, 2002

[26] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, "Characterization of Four-Class Motor Imagery EEG Data for the BCI-competition 2005," *Journal of Neural Engineering*, vol. 2, no. 4, pp. 14-22, Dec. 2005.

[27] B. Blankertz, BCI Competition III Webpage. [Online]. Available: http://ida.first.fraunhofer.de/projects/bci/competition˙ii˙9

[28] L.A. Farwell, E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography & Clinical Neurophysiology*, vol. 70, no. 6, pp. 510-523, Dec. 1988.

[29] B. Blankertz, BCI Competition II (2003) – P300 Speller Dataset (webpage). Available online at: http://www.bbci.de/competition/ii/, Documentation: http://www.bbci.de/competition/ii/albany_desc/albany_desc_ii.pdf

[30] Ioana Adam, C. Nafornita, J.-M Boucher, A. Isar, "A Bayesian Approach of Hyperanalytic Wavelet Transform Based Denoising," *International Symposium on IEEE, Intelligent Signal Processing*, pp. 1-6, Oct. 3-5, 2007.

[31] John L Semmlow, *Biosignal and biomedical image processing: MATLAB-based applications*. Vol. 22. CRC press, 2004.

[32] Elif Derya Übeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification," *Digital Signal Processing*, vol.19, no.2, pp. 297-308, Mar 2009.

[33] Xiaolan Zhao, Wu Zuguo , Jiren Xu, Keren Wang, "Speech Signal Feature Extraction Based on Wavelet Transform", *IEEE Conference on Intelligent Computation and Bio-Medical Instrumentation*, pp.179-182, Dec. 14-17,2011.

[34] Inan Güler and Elif Derya Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients" *Journal of neuroscience methods*, vol. 148, no. 2, pp. 113-121, Apr 2005.

[35] Hyekyoung Lee and Choi Seungjin, "PCA-based linear dynamical systems for multichannel EEG classification" *Proceedings of the 9th International Conference on. Neural Information Processing*, vol. 2, Nov. 18-22, 2002.

[36] Ingrid Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961-1005, Sep. 1990.

[37] Dipti Patra, Manab Kumar Das, and Smita Pradhan, "Integration of FCM, PCA and Neural networks for classification of ECG Arrhythmias," *IAENG International Journal of Computer Science.* vol. 36, no. 3, pp. 24-62, Feb 2010.

[38] Travis C.Collier and Charles Taylor, "Self-organization in sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 866-873, July 2004.

[39] Shang-Ming Zhou, Q. Gan John and Francisco Sepulveda, “Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface,” *Information Sciences*, vol. 178, no. 6, pp.1629-1640, Mar. 2008.

[40] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal, “Introduction to multi-layer feed-forward neural networks”, *Chemometrics and intelligent laboratory systems*, vol. 39, no. 1, pp.43-62, Nov. 1997.

[41] Elif Derya Übeyli, “Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals,” *Expert Systems with Applications*, vol. 37, no. 1, pp. 233-239, Jan. 2010.

[42] Burges, JC. Christopher, “A tutorial on support vector machines for pattern recognition,” *Data mining and knowledge discovery*. vol. 2, no. 2, pp. 121-167, 1998.

[43] Simon S. Haykin, *Neural networks and learning machines*. New Delhi: Pearson Education Asia, 2002.

[44] Corinna Cortes and Vladimir Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273-297, 1995.

[45] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, H. Ritter, “BCI Competition 2003 – Dataset IIb: Support Vector Machines for the P300 Speller Paradigm,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1073 – 1076, June 2004.

[46] V. Asadpour, M. Vatankhah, and Reza Fazel-Rezaie, “Perceptual Pain Classification using ANFIS adapted RBF Kernel Support Vector Machine for Therapeutic Usage,” *Applied Soft Computing*, vol. 13, no. 5, pp. 2537–2546, May 2013.

[47] Yongmei Michelle Wang, T. Schultz Robert, R. Todd Constable, Lawrence H. Staib, “Nonlinear estimation and modeling of fMRI data using spatio-temporal support vector regression,” *Information Processing in Medical Imaging Springer Berlin Heidelberg*, pp. 647-659, 2003.