Motor Imagery based EEG Signal Classification

Submitted by:

Muhammad Zeeshan Baig

(2011-NUST-MS-RIME-10)

Thesis Supervisor:

Dr. Yasar Ayaz

MS

Robotics & Intelligent Machine Engineering

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING (SMME),
NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY (NUST),
SECTOR H-12, ISLAMABAD, PAKISTAN

May 2015

Dedication

I would like to dedicate my research work to my beloved parents,

All My Friends

&

Family Members

whom continuous encouragement, guidance and support remained a true source of motivation for me. I have always looked upon them, as a real source of Morale booster.

Also, to my knowledgeable Instructor Dr. Yasar Ayaz, who enacted like a beacon light.

Certificate of Originality

The substance of this thesis is the original work of the author and due reference and acknowledgement has been made, where necessary, to the work of others. No part of this thesis has been already accepted for any degree, and it is not being currently submitted in candidature of any degree.

Muhammad Zeeshan Baig

2011-MS-NUST-RIME-10

Thesis Scholar

Countersigned

Dr. Yasar Ayaz

Thesis Supervisor

Acknowledgement

All regards to Almighty Allah, the most beneficial, the most merciful and creator of everything what we know and what we don't.

The current work is an absolute example of the team effort and support offered and rendered by the faculty and staff members of School of Mechanical and Manufacturing Engineering (SMME), NUST.

The contributions and contemporary technical support offered by Dr. Yasar Ayaz (HOD Robotics & Intelligent Machine Engineering) really deserve special recognition and acknowledgement.

An exemplary valuable guidance and support for the wholesome project has been provided by the Guidance Committee members, Assistant Professor Dr. Omer Gilani, Dr. Mohsin Jamil and Dr. Muhammad Naveed.

I would like to thank my family in general and my father, mother for their consistence moral support and sparing their precious share of my time.

Lastly, I sincerely thank to Engr Waseem Afzal, Engr. Shahwar Yasin, Engr. Badar Ali and Engr. Shehzad Ahmed, whom assistance and support throughout the tenure, made the transformation of an elusive project plan to current work reality.

Muhammad Zeeshan Baig

Table of Contents

List of Figures

Abstract

Effective classification of Motor Imagery (MI) tasks based EEG signals is the main hurdle in order to develop online Brain Computer interface (BCI) system. The key part of BCI system is to extract the dominant features from EEG data along with selection of a suitable classifier.

In this research thesis, a relatively new approach has been implemented to accurately classify EEG signals that have been extracted from MI. The data-set was obtained from BCI competition-II 2003 named Graz database. Two channels have been selected for preprocessing i.e. C3 and C4. After applying pre-processing techniques feature vector have been extracted. The feature vector consists of bior Wavelet Transform (WT) coefficients, Power Spectral Density (PSD) approximations, average power and aggregated EEG signal. In this study, we have presented a comparison of mostly used classification algorithm with relatively new classification technique i.e. Self-organizing maps (SOM) and Deep Belief Nets (DBN). It has been depicted from measured data that SOM shows a classification of 84.17% on Principal Component Analysis (PCA) implemented reduce dataset. Furthermore, a 2% increase in classification accuracy has been attained by using bi-orthogonal filter banks for wavelet transform instead of Daubechies WT.

In Deep Learning, Firstly a weak classifier has been trained using deep belief networks (DBN) after that the concept of boosting has been applied in order to make the classifier strong. The boosting algorithm that has been implemented in this research is ada boosting. Multilayered structure has been used for DBN consisting of hidden units and hidden layers. Furthermore, the performance has also been tested using different hidden units and hidden layers. The experimental results shows that with different hidden layer there is a significant change in classification results but overall performance is better for 15 hidden layers network. The results are compared with different state of the art classification algorithms i.e. Support Vector Machine (SVM) and Self organizing maps (SOM) based classification techniques and DBN shows better results with a recognition error of minimum 6% in the classification performance.

Chapter 1

INTRODUCTION

1.1 Introduction

One of the main purposes of our existence on Earth is to Interact with other human beings. However, in some unforeseen cases the normal way of interaction cannot be established by all due to some diseases or accidents. For such peoples (also known as disabled ones or locked-in) Brain Computer Interfaces (BCIs) provides an alternate method to sense and show the feelings to external environment. Researchers all around the globe linked to different fields of study are striving hard to improve the lifestyle of suffered peoples. Most of the research has been carried out to improve and enhance the accuracy and real time speed of BCIs systems by establishing state of the art algorithms to acquire signals, analyze brain activities and improve efficient classification algorithms to develop online BCIs.

The BCI is a device that permit brain signals to interact with the environment. BCI has been categorized into two types namely, invasive BCI and non-invasive BCI [1]. In invasive BCI, the electrodes are mounted in to the brain skin to extract signals (require surgery) and in non-invasive BCI the electrodes are mounted on the surface of the scalp to acquire the signals. BCI system has been used to help paralysis, quadriplegics and amyotrophic lateral sclerosis people to drive computers and machines directly by brain signals rather than by physical means and it is equally useful for non-disable individuals [2]. BCI system can also be applied in different areas included robotics, biomedical technologies, surgery etc. [1], [2].

There are many sources to measure brain activities for BCI.

Following signals have been used to power up the invasive BCIs, where the electrodes are planted into the skin to measure the brain activities.

- electrocorticograpy (ECoG)
- single micro-electrode (ME)

- micro-electrode array (MEA)
- local field potentials (LFPs)

In non-invasive BCIs, the signals used to drive the systems can be classified as:

- electroencephalography (EEG)
- magnetoencephalography (MEG)
- Functional Magnetic Resonance Imaging (fMRI)
- Near Infrared Spectroscopy (NIRS)

The BCI system with EEG input has been the most reliable and frequently used source to measure brain activity due to the non-invasive EEG electrodes availability, low hardware cost and transferability. It also exhibit high temporal resolution [1], [2], [3].

Acquiring of EEG signals for BCI can be done in different ways that are very useful. Some of the methods required an event to generate the EEG signals and some are event independent. Motor Imagery (MI) is one of the methods used to generate EEG signals that is related to motor movements is studied in this article. MI based EEG signals have been applied to many BCIs application where these signals have been controlled to open the interface with the external environments [4].

To convert the MI based EEG signals to BCI input decision and control signals different techniques and algorithms have been used. These algorithms and techniques can be defined in three steps.

1. Signal Enhancement
2. Feature Extraction
3. Classification

In signal enhancement which is also known as signal pre-processing technique, different filters are used in order to reduce noise and amplify the information in the signal. In 2nd step the data is transformed into the features that give the maximum information about the brain signal. The feature extraction part is based on the signal acquisition and application. The last step is to classify the signals based on some mathematical model using the features extracted in 2^{nd} step. The task of mathematical model is to generate the control signals specific to the application.

1.2 Aim and Objective

In this work, the EEG signal has been enhanced by applying Band-Pass Filter, Median filter and variant of Laplacian filters. Common Spatial Pattern (CSP), Wavelet Transform (WT) and Power Spectral Density have been used to extract the required features out of EEG signals. Principal Component Analysis (PCA) has been used to reduce the size of feature set. In addition to that further normalization techniques have also been applied to extract features. Classification has been carried by implementing efficient classification algorithms like Support Vector Machines (SVM), k- Nearest Neighbor (k-NN), Linear Discriminant Analysis (LDA), Quadrature Discriminant Analysis (QDA), Self-Organizing Maps (SOM) based Neural Networks and Deep Belief Networks (DBN). Database used for evaluating these methods is from BCI Competition III named Graz database.

The main objective of this writing is to show a different classification approach i.e. Self-organizing map based neural network and deep learning based classification with a comparison to the other classification algorithms. Classification is performed both on original features as well as reduced features extracted from raw EEG signals.

1.3 Thesis Outline

The first chapter of thesis contain introduction followed by BCI research and development in Chapter 2. Chapter 3 discusses the Signal Enhancement techniques used in this study. After that the Results of classification algorithms has been written in Chapter 4 with simulation results. Chapter 5 will show the overall summary and conclusion of this writing Based on results.

Chapter 2

LITERATURE REVIEW

The device that translate brain signal into control and decision signals to interact and control is commonly known as Brain Computer Interfaces (BCIs). The BCI system can serve as the preferred pathway for

Figure 1 Severe diseases related to Brain [5]

For that reason, many scientist and researchers related to different areas have been providing their input in this matter to advance the class of living of the disabled ones.

The main parts of a typical BCI system related to acquisition and operations are shown in Figure 2. In this Chapter, Every part will be defined with brief explanation of working. First we are going to explain the signal acquisition part. Then analysis of the acquired signals will be given and how this analysis is related to the BCI system application is elaborated. After that signal processing and classification is going to be discussed followed by some of the examples of the BCI system reported in literature.

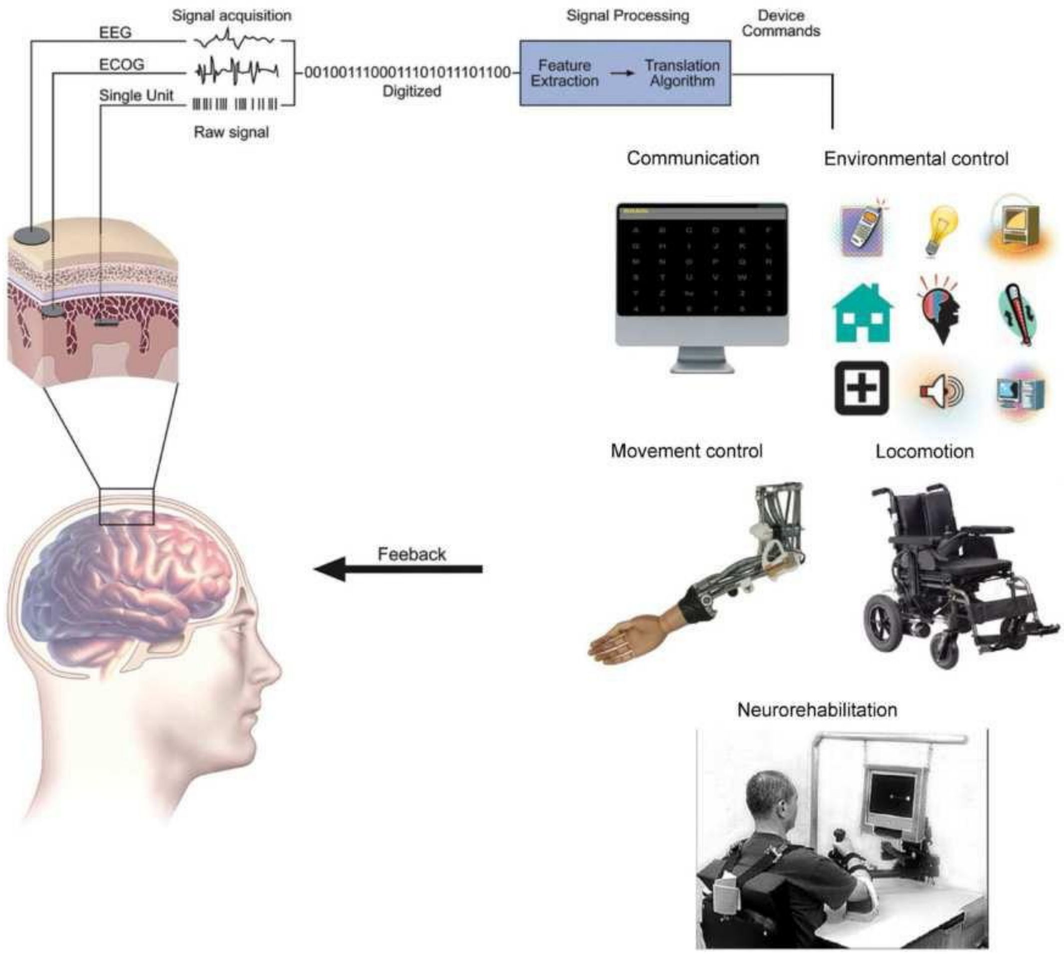

Figure 2 the BCI Development Cycle [6]

2.1 BCI Signal Acquisition

Signal acquisition part for BCI device is categorized in to two classes. First one is non-invasive, in which the activity of brain is recorded by placing electrodes on surface of head. EEG and MEG are the examples of non-invasive brain signal acquisition. The second type is invasive, in which electrodes are placed into the skull in order to get the required brain activity. The examples of invasive acquisition system are ECoG, ME, MEA, and LFPs. There are other methods used to power up the BCI system which includes Functional Magnetic Resonance Imaging (fMRI) and Near Infrared Spectroscopy (NIRS) which are also non-invasive [1], [2]. A generic graph between temporal resolution and spatial resolution of different signal acquisition techniques has been shown in Figure 3.

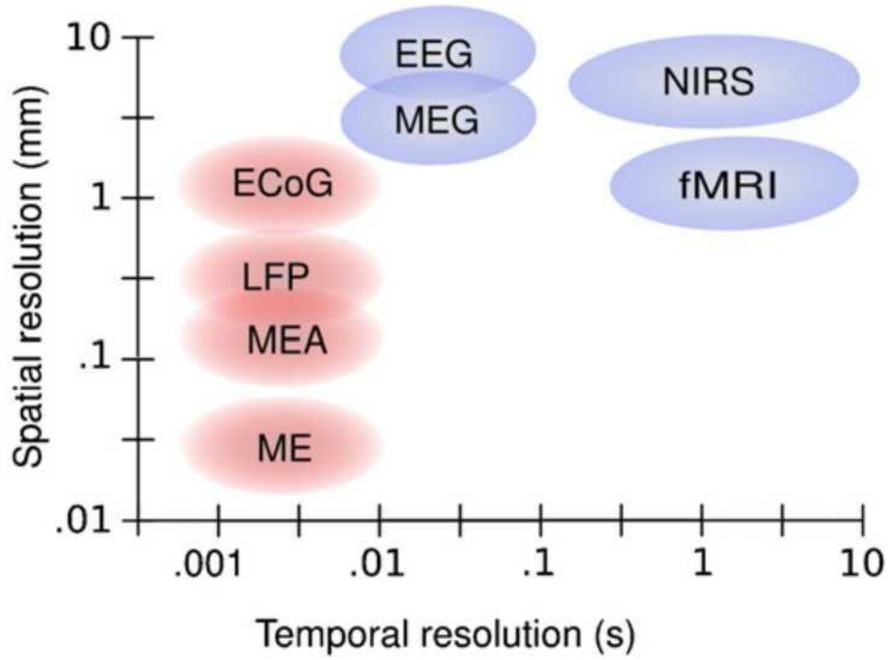

Figure 3 Comparison of Spatial and Temporal Resolution [7]

2.2 Electroencephalography (EEG)

The electrical activity of brain measured through electrodes that are placed on the brain skin is known as Electroencephalography (EEG) is [8], [9]. It is first recorded by Hans Berger in 1929. The EEG signal is of very small amplitude mostly of the order 10-4 Volts which has some noise. The EEG signal is then amplified to increase information strength and filtered to extract the desired frequency and also to suppress noise. After amplification and filtering the signal is digitized and recorded (Figure 4).

Figure 4 EEG signal acquisition system including electrodes cap, an amplifier and a monitor to see recorded signals [10]

Main reason to use EEG is because of its high temporal resolution that is effective in online applications, but its resolution related to spatial domain is quite low. The low spatial resolution is due to the blurring phenomena of brain tissues, the acquired EEG signals have some artifacts which generates due to eye movement, deviation from actual position of electrodes and also from body movements. Additionally, in non-invasive acquisition EEG electrodes are positioned on the surface of the brain by rub in a conducting gel to enhance skin conductance. This also contributes towards the negative side in terms of practical use. After all these disadvantages EEG acquisition system to record brain signal is the most extensively used input for power up BCI systems because of its low cost and portability. [3].

2.2.1 Motor Imagery based EEG Signal

Motor imagery (MI) is a psychological procedure in which a person practice or execute a given motor task. MI based EEG signals are extensively used in sport training as cerebral exercise of motion, nervous restoration, and has also been implemented as an investigation model in cerebral neuroscience and cognitive psychology to examine the data.[11][12]. The activation in motor imagery part of brain has been shown in Figure 5.

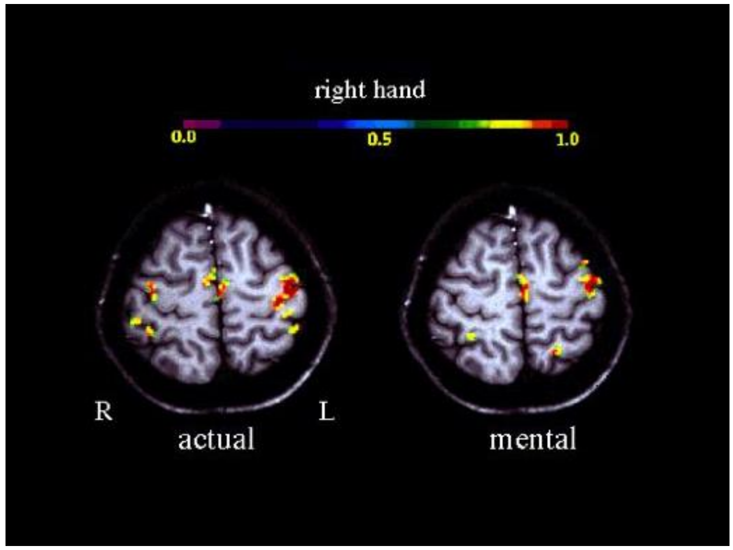

Figure 5 Initiation in motor cortex in the course of motor imagery [11]

An example of motor imagery based EEG signal is shown in Figure 6 and Figure 7.

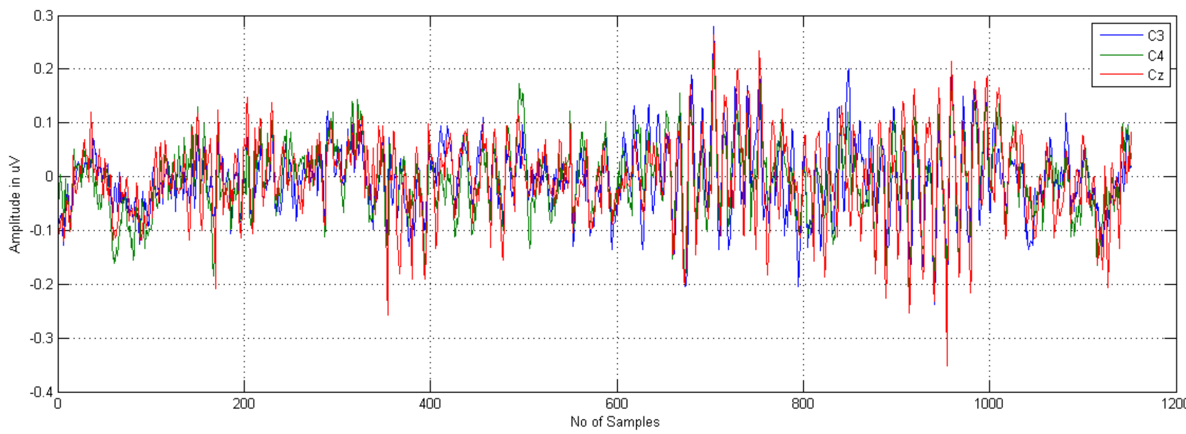

Figure 6 Motor Imagery based EEG signals for Left Hand Movement

Figure 6 contains the EEG Signals generated by motor imagery based task in which subject moves his left hand to generate signals, similarly Figure 7 shows the same EEG signals generated by the movement of right hand.

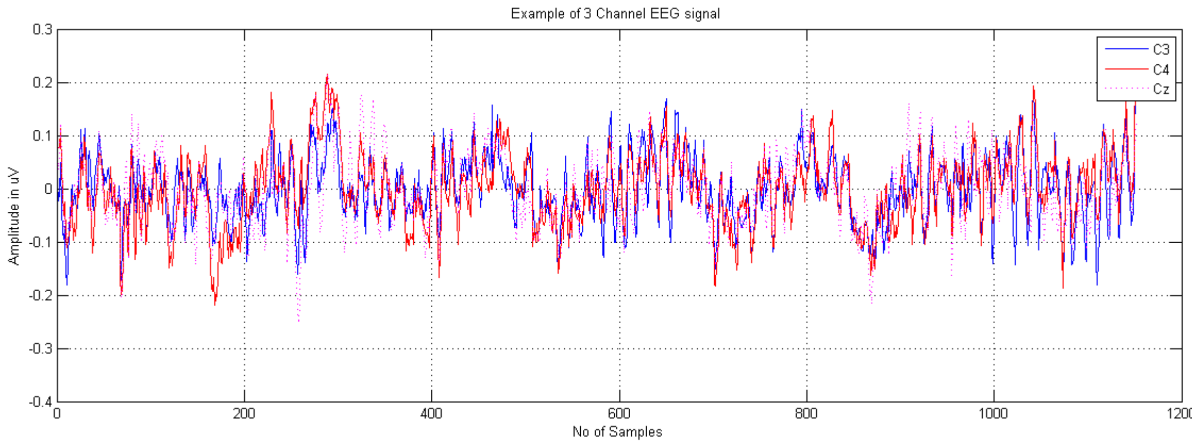

Figure 7 Motor Imagery EEG signals of Electrodes C3, C4 and Cz for Right Hand Movement

2.3 Signal Processing

To convert the MI based EEG signals to BCI input decision and control signals different techniques and algorithms have been used which comes under the umbrella of signal processing. These algorithms and techniques can be defined in three steps.

1. Signal Enhancement
2. Feature Extraction
3. Classification

In signal enhancement which is also known as signal pre-processing technique, different filters are used in order to reduce noise and amplify the information in the signal by implementing techniques and algorithms like filtering the signal, down-sampling original signal, etc. In 2nd step the data is transformed in to features that give the maximum statistics about the brain signal. The feature extraction part is based on the signal acquisition and application. The last step is to classify the signals based on some mathematical model using the features extracted in 2nd step. The task of mathematical model is to generate the control signals specific to the application. Detailed information on techniques used in research has been given in Section 3.3.

2.4 Applications of Brain Computer Interface

In this segment, some of the application related to Brain Computer Interface including Communication, Environmental control, Movement control and Locomotion has been discussed.

2.4.1 Communication

P300 Spelling paradigm has been used wieldy for communication starting from letter, words to sentences and researchers have achieved an accuracy of 95% for P300 Speller. Further information related to P300 speller is available at [13]. A simple stimulus for P300 Speller has been shown in Figure 8.

Figure 8 P300 speller matrix of size 6 x 6. The white row is intensified [17].

2.4.2 Environmental Control

One application of BCI system for person with disabilities is to control different devices. A system was designed, implemented and tested on a person who is disable to communicate with the devices in near environment. [14].

2.4.3 Movement Control

Different researchers around the globe are working on rehabilitation of motor control with some prosthetic or robotic device to develop a system for paralyzed patients.

2.4.4 Locomotion

There is an advanced research carried out to develop a BCI driven wheel chair to improve the standard of living for paralyzed people. A stimulus used in [15, 16] has been shown below in Figure 9 and 10.

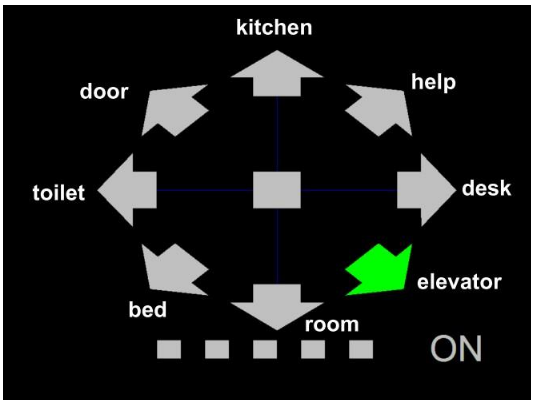

Figure 9 P300 paradigms used for locomotion [15]

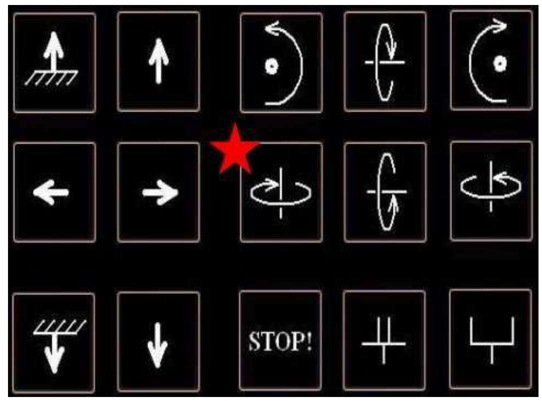

Figure 10 P300 paradigms used for locomotion [16]

Chapter 3

DIGITAL SIGNAL PROCESSING IN BCI

In Chapter 2, a short-term explanation of overall signal processing process in a classic Brain Computer Interface system is given. In this chapter, the methods will be explained in details and main focus will remain on those techniques that has used in the study. The summary of overall signal process can be given as:

- For signal enhancement: for extracting required signal band-pass filtering has been used, Median filtering has been used to remove pulse noise.
- For extracting features: Wavelet Transforms (WT), Power Spectral Density (PSD), Aggregated EEG Signal and Principal Component Analysis (PCA).
- For classification: Support Vector Machines (SVM), Deep Belief Nets (DBN), Self-Organizing Maps (SOM)
- For evaluation: classification accuracy, Mini Batch Mean Square Error, Full Batch Mean Square Error.

3.1 Signal Enhancement

EEG signals has a low spatial resolution, compare to its temporal resolution, which is due to the blurring phenomena caused by the mass conduction in intervening tissues and also the EEG signal contain some artifacts that belongs to eye, electrodes and muscular movement, So in result EEG signals has comparatively low signal to noise ratio. To improve signal quality some kind of signal enhancement mechanism is needed. In this study, two enhancement techniques has been implemented, band pass filtering and median filtering.

3.1.1 Band Pass Filtering

To extract the frequency band of 0.5 - 30 HZ, a Butterworth filter has been designed in MATLAB. The filter of order 64 and sampling rate of 128 Hz has been designed to extract rhythms (α and β bands; 8-12 Hz and 13-30 Hz, respectively) associated with sensorimotor events [21]. The equation for Butterworth filter is given as.

$$G(w) = \sqrt{1/(1 + w^{2n}}$$

3.1.2 Median Filtering

The mostly used nonlinear filtering method is median filtering which efficiently degrade the interference pulse while maintain the original characteristic of the signals. That's why it is widely used as a preprocessing technique. The length of the filtering window is describe as n where signal length is N. the output of the filter is given by the function

$$med(a_i) = \begin{cases} a_{k+1} & n = 2k + 1\ (Odd) \\ \dfrac{[a_k + a_{k+1}]}{2} & n = 2k\ (even) \end{cases}$$

Here a_k is the $k - th$ maximum observed data and $a_1, a_2, a_3 \ldots a_k$are the observed data. Consider an example in which dataset contains 7 samples i.e. {2, 3.5, 1, 3, 1.5, 4, 2.5} then output of median filter is 2.5. the signal will remain as it is if the pulse has a length of k+1 or greater else it will be degraded from the sequence and it is the highlighting characteristic of the median filter that it eliminate the pulse noise and local details remain intact. After this technique the resultant signal is then provided to the feature extraction block where the wavelet transform is applied to the signals to extract features.

In order to remove pulse noise median filter is applied with a window size of 50 and length of a single trail is 769. The median filter is regulating with zero mean and unity variance. The results of median filtering on left hand movement has shown in Figure 11 and for right hand movement is recorded in Figure 12.

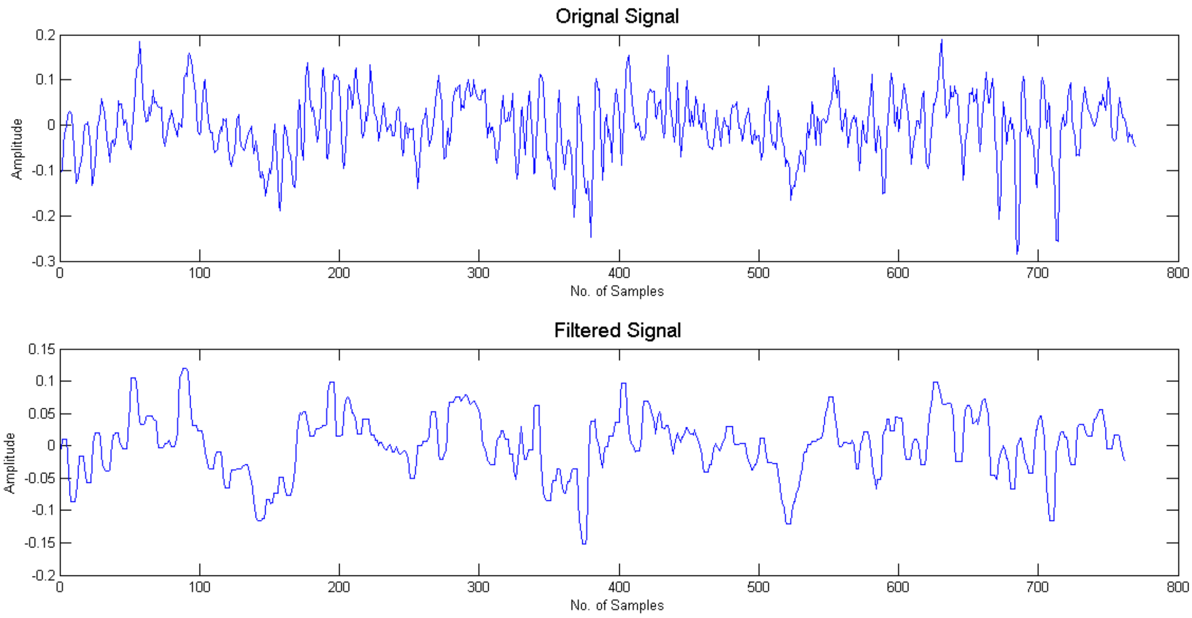

Figure 11 Noise removal using median filtering (Left hand movement)

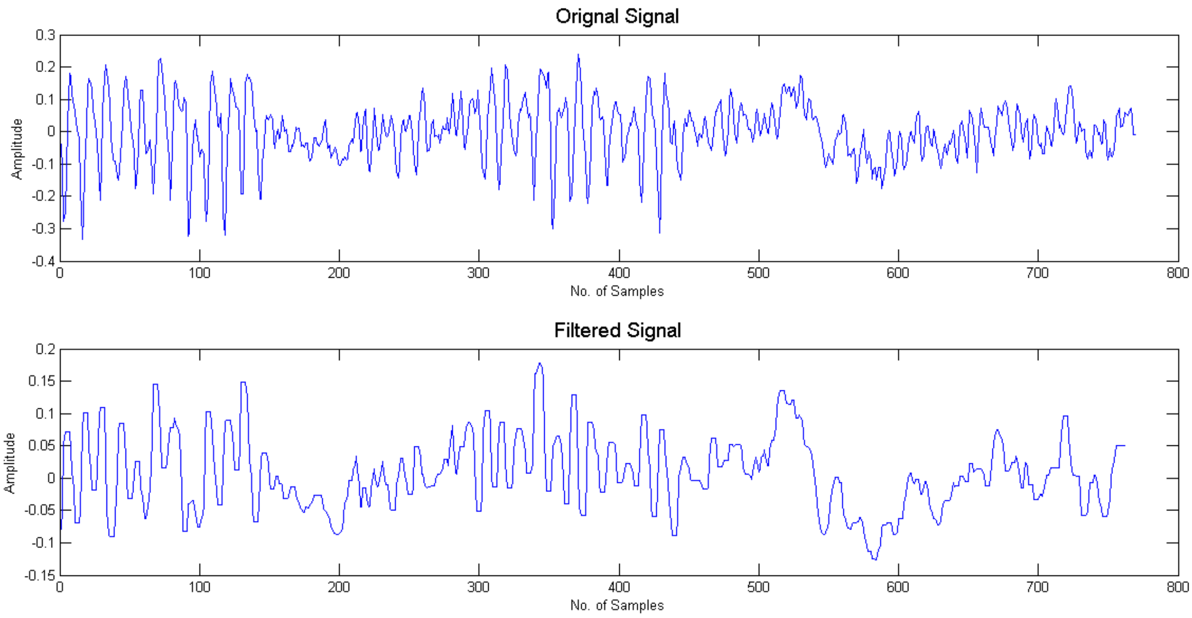

Figure 12 Noise removal using median filtering (Right hand movement)

3.2 Feature Extraction

The part of extracting the desired information from raw EEG signal requires some techniques known as feature extraction. In BCI several feature extraction techniques has been used which is quite related to the application and type of the EEG signal based on generation. We have extracted different features for motor imagery based EEG signals given below:

Feature vectors have been extracted from the predefined channels C3 and C4 [18]. The feature vector based on WT and statistical parameters of the selected EEG channels has been used by saugat in [19] with a little modification in using wavelet transform. We have used the same features In order to compare the predefine techniques.

3.2.1 Wavelet transform

The inability to tackle non-stationary signals has been the main hurdle in the Fourier transform (FT) as it neglect the small changes in high frequency components [20]. On other hand Wavelet transform (WT) has capability to distinguish spatial domain features of a signal from temporal features, that's why WT has an upper hand over the FT while extracting the features. EEG signals from C3 and C4 has been decomposed through a bi-orthogonal Wavelet transform rather than Daubechies Wavelet Transform [19] to acquire the frequency bands signals.

The wavelet function $\psi(t)\epsilon L^2(R)$ has zero mean

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \qquad (1)$$

The mother wavelet is given by

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)_{u,s\epsilon R,u>0} \qquad (2)$$

Where µ the scattering parameter, s is the scaling parameter and R defines the wavelet space. In this article bi-orthogonal 6.8 (bior6.8) mother wavelet transform has been used to extract the frequency band as shown in Table.1.

Table 1 Frequency Band of EEG signals

Delta	[0 – 4 Hz]
Theta	[4 – 8 Hz]
Alpha	[8 – 13 Hz]
Beta	[13 – 30 Hz]

The wavelet and scaling function of bior6.8 is shown in Figure 13 and Figure 14.

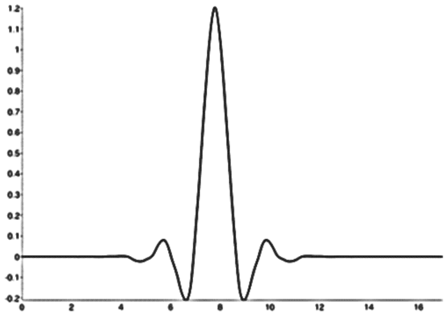

Figure 13 Decomposition scaling function φ

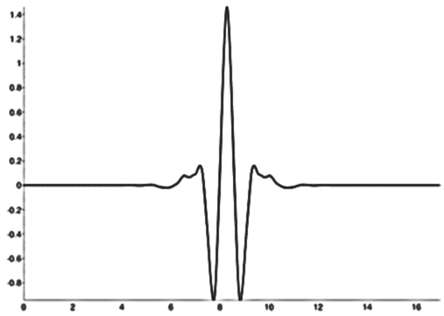

Figure 14 Decomposition wavelet function ψ

The bior 6.8 wavelet transform of level 5 has been shown in Figure 15

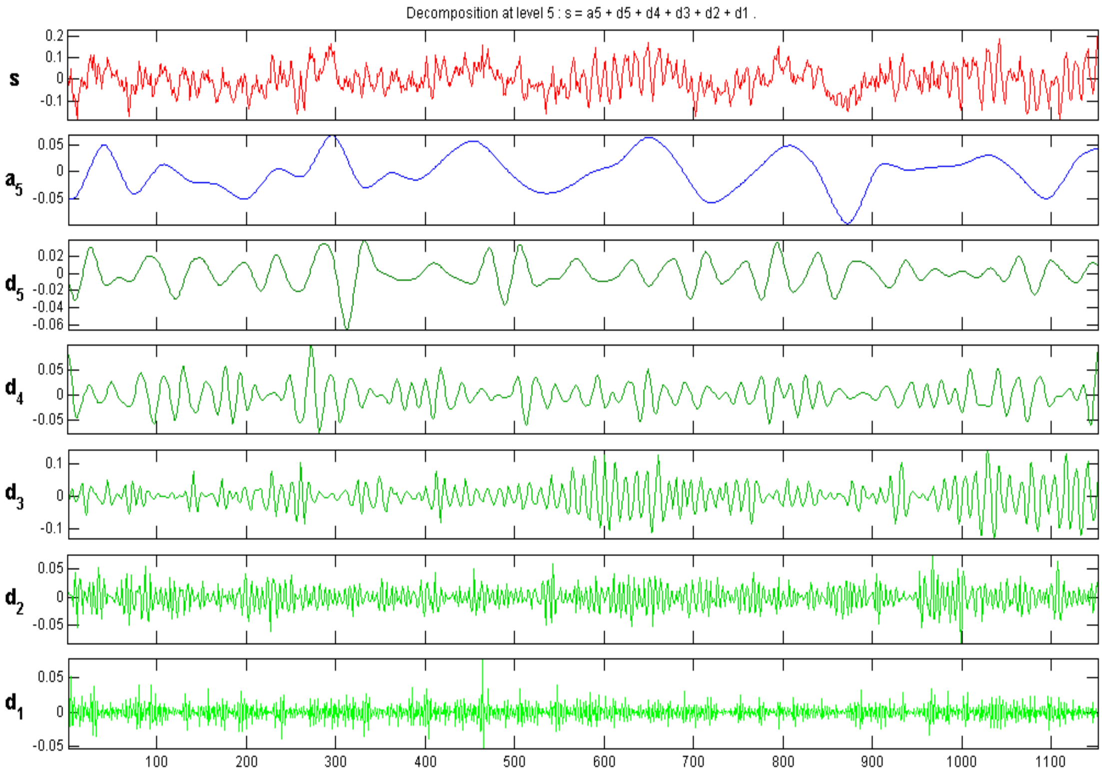

Figure 15 Wavelet Transform of level 5 using Bior 6.8

The wavelet coefficient d4 of bior6.8 wavelet transform has been selected as a feature shown in Figure 16 and Figure 17.

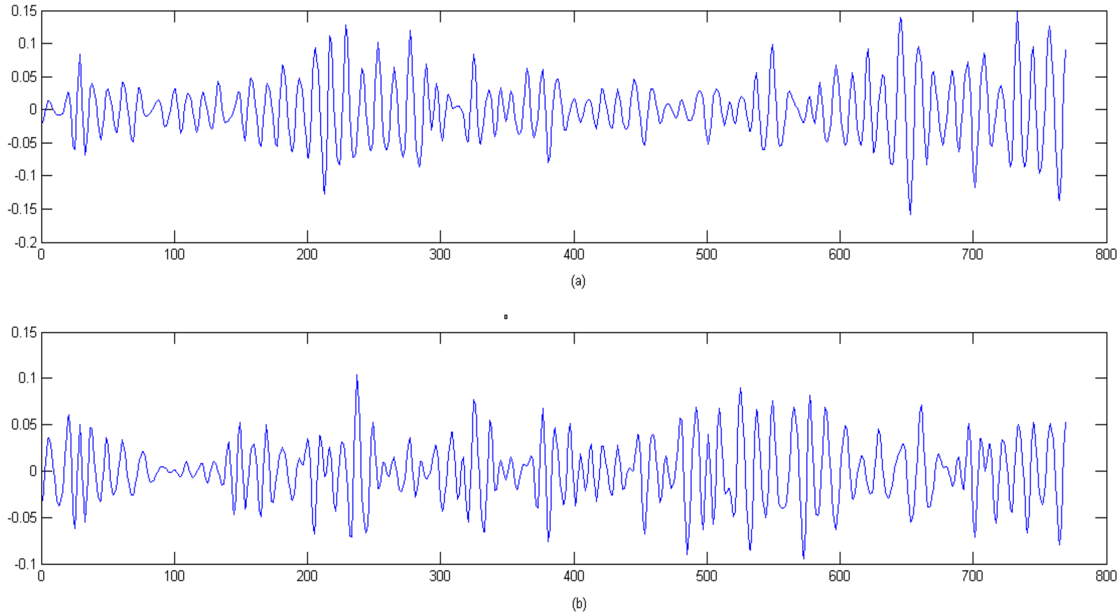

Figure 16 Wavelet coefficients Of Left signal (a) electrode C3 (b) electrode C4

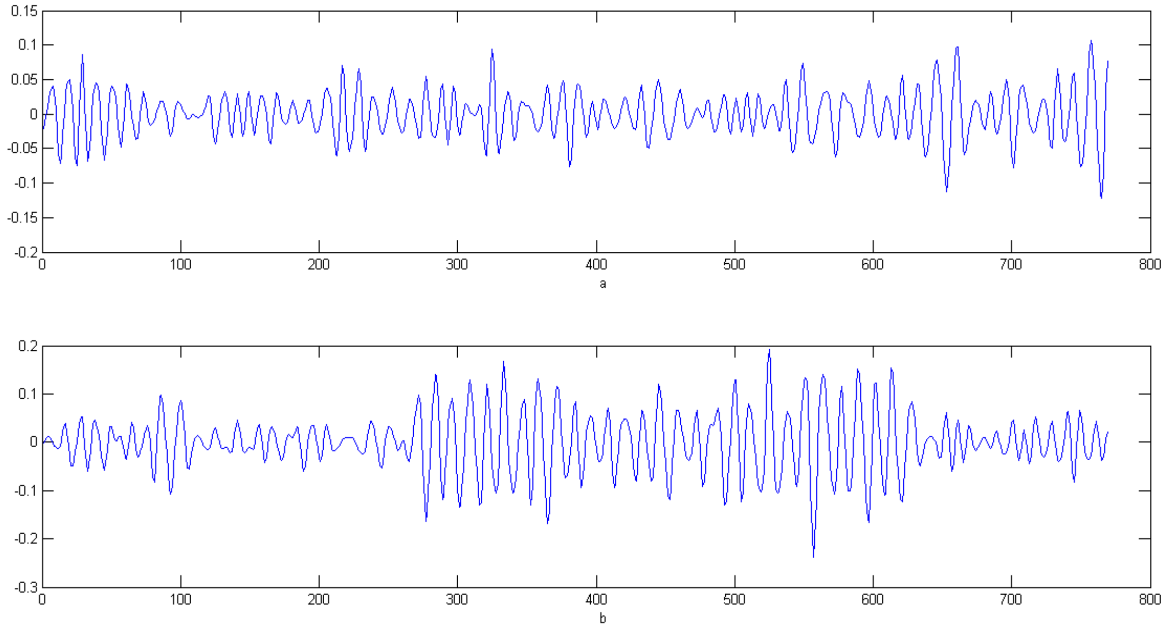

Figure 17 Wavelet coefficients of Right Hand signal (a) electrode C3 (b) electrode C4

3.2.2 Spectral Estimation Method

Power Spectral density (PSD) has been used to extract the signal information in order to have the knowledge of frequency vs. power spreading. PSD is the autocorrelation of Fourier transform (FT) that has been considered stationary in a wide range [21]. So this has been a good approach to segment out complete data for an EEG signal. The Welch PSD estimate has been carried out with a Hamming window of 64 [19]. To compute the periodogram of overlapping segments a Welch method has been used that splits input into overlying pieces and then the PSD approximations has been calculated which is the average of that data. The PSD of EEG signal has been shown in Figure 18.

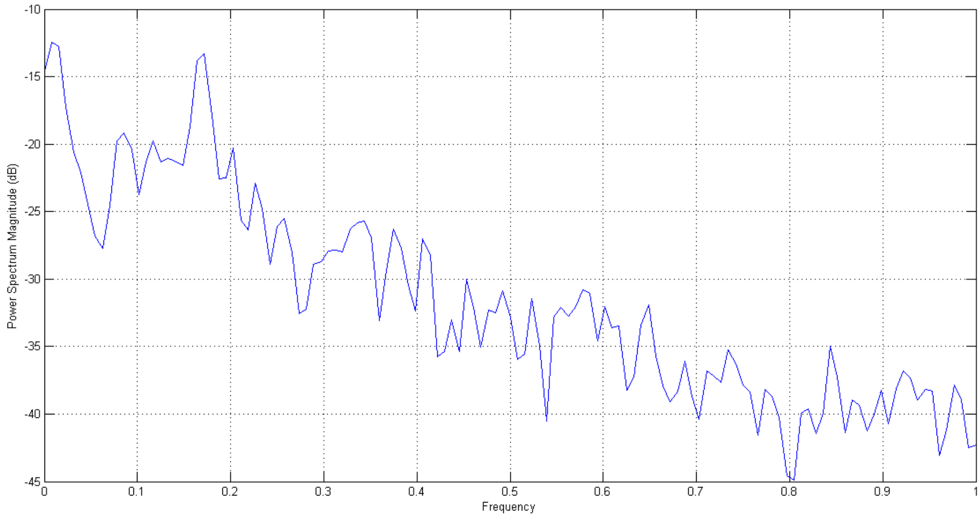

Figure 18 PSD of EEG Signal

The PSD estimates 8-25 Hz has been extracted in which 8-12Hz correspond to α and μ band and 18-25Hz correspond to the β band. Mean power has also been computed for each band.

3.2.3 Aggregated EEG signal

A new technique of aggregating EEG signal to generate a feature set has been implemented in this research. We aggregate the data of three channel C3, C4 and Cz by taking absolute of the signals and combine the signals using a time window of 2 seconds. By aggregating the channels we enhance the information related to motor imager of the three channels that occurs for the given time span.

3.2.4 Principal Component Analysis (PCA)

In order to reduce the size of feature set Principal component analysis (PCA) has been used and it is a conventional statistical algorithm to reduce dimensionality developed by Pearson K. in 1901 [22]. This method has been broadly used in data compression, dimensionality reduction and data analysis to transform a set of samples of probably associated variables into a set of samples that are uncorrelated known as principal components. PCA is used to reduce the size of feature set in this work.

3.3 Classification

Classification is to classify the signal feature sets into their particular classes with maximum accuracy. This can be accomplished in two ways, first one is unsupervised learning in which no information about the class label has been given. The second is supervised learning in which every sample in the dataset has a class label and it is divided in to two set training and testing datasets, training dataset has been used to train the classifier and testing dataset has been used to evaluate the classifier. [23]. In BCI study, Supervised learning is preferred over unsupervised learning.

In order to do so we used different classifiers and compared the results of Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Linear Support Vector Machine (SVM) and k-nearest neighbor (kNN), SOM based neural networks and Deep Belief Nets (DBN).

3.3.1 Linear Discriminant Analysis

LDA's main functionality is to provide the feature's spread of two sets normal with similar covariance matrix [24]. LDA reduce the dimensionality by projecting multidimensional data into a line reducing L spreading to (L-1) dimensional spreading. LDA provides maximal separability by enhancing the ratio of between-class variance and within-class variance. Figure 19 shows the example of LDA.

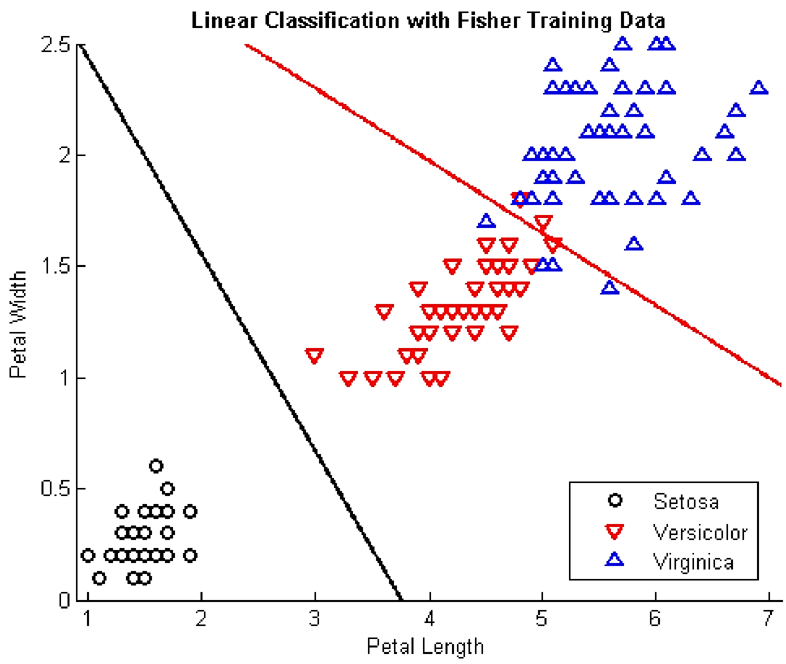

Figure 19 Example of LDA

3.3.2 Quadratic Discriminant Analysis (QDA)

In comparison, QDA is a comprehensive form of LDA, on condition of two classes and the groups are normally dispersed [24].On the other hand in contrast to LDA, QDA didn't give attention to covariance of classes. The surface divides the low dimensional space will be a conic section (like circle, parabola, etc.). The example of QDA is shown in Figure 20.

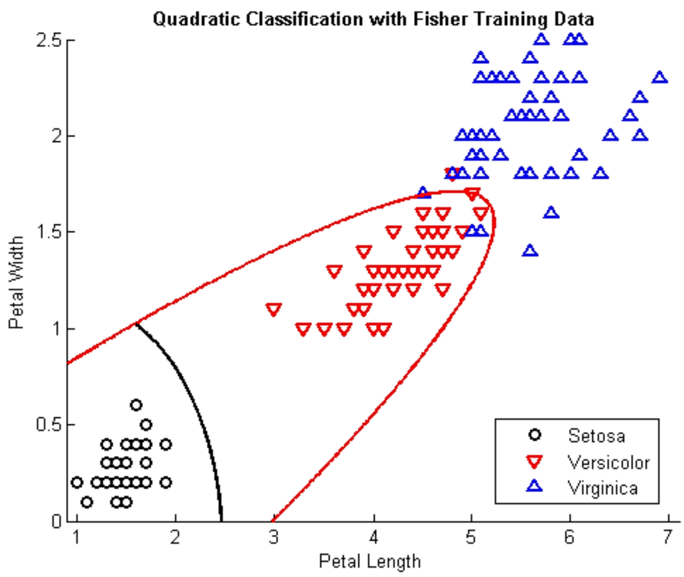

Figure 20 Example of QDA

3.3.3 Linear Support vector machine (LSVM)

In supervised learning techniques SVMs are very popular for classification. As SVM is generalized linear classifiers, so it can directly applied on both untransformed and non-linear transformed feature sets [25, 26]. SVM makes a maximal dividing hyper plane with a maximum threshold amongst the groups; by increasing the dimensionality of feature space as depicted in Figure 21.

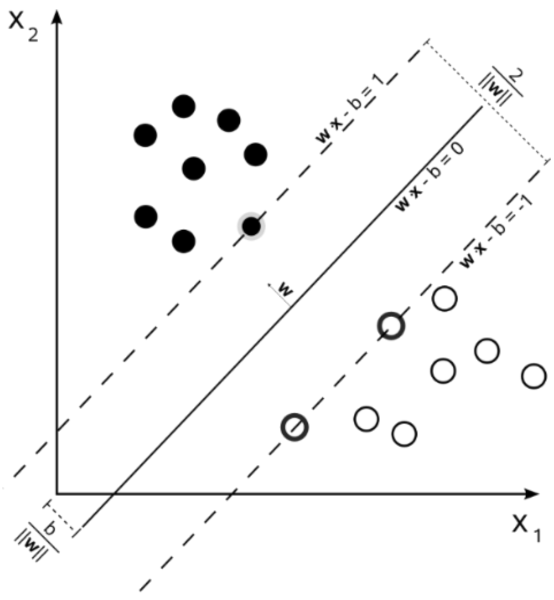

Figure 21 SVM Example

Consider a training set X defined as $\{x_i, i=1, 2 \ldots.., n\}$ belongs to one of the two classes ω_1 and ω_2 with corresponding labels $y_i=\pm1$. The function $\gamma(x)=\omega^T x+\omega_o$ is known to be the discriminant function where, ω is the weight of coefficient vector, and ω_0 defines the threshold. Classifying rule is

$$\omega^T x + \omega_0 > 0 \Rightarrow x \in \omega_1; y_i = +1$$
$$\omega^T x + \omega_0 < 0 \Rightarrow x \in \omega_2; y_i = -1$$

A margin b (b>0) is introduced, so that the solution becomes

$$y_i(\omega^T x + \omega_0) \geq b$$

Where the points whose distance is greater than b form the dividing hyper plane. If b=l, the canonical hyper planes (H1 and H2) are given by:

$$H_1: \omega^T x + \omega_0 = +1$$
$$H_2: \omega^T x + \omega_0 = -1$$

Thus we have,

$$\omega^T x + \omega_0 \geq +1; for\ y_i = +1$$
$$\omega^T x + \omega_0 \geq -1; for\ y_i = -1$$

3.3.4 k - Nearest Neighbour (kNN)

The main difference of kNN algorithm is decision making in order to create the training dataset more generalize until a query or data is came across that is not seen before. The basic supposition in kNN is of making class probabilities almost constant for a set that's make kNN simplest among all machine learning technique. In order to classify, the kNN algorithm discover the k-closest neighbors in training dataset, where the classes of closest neighbors are used to evaluate the class nominees. K is normally a small non-negative integer. The mostly used methods to compute distance are, Manhattan distance, Mahalanobis distance and Euclidean distance. Two factors that affect the performance of the algorithm are: an appropriate match function and a proper k. Figure 33. Shows an example of kNN classification algorithm. If k is very large then there will be overlapping of large and small classes and if the value of k is very small, then no improvement of k-Nearest Neighbour classification algorithm is outlined [27].

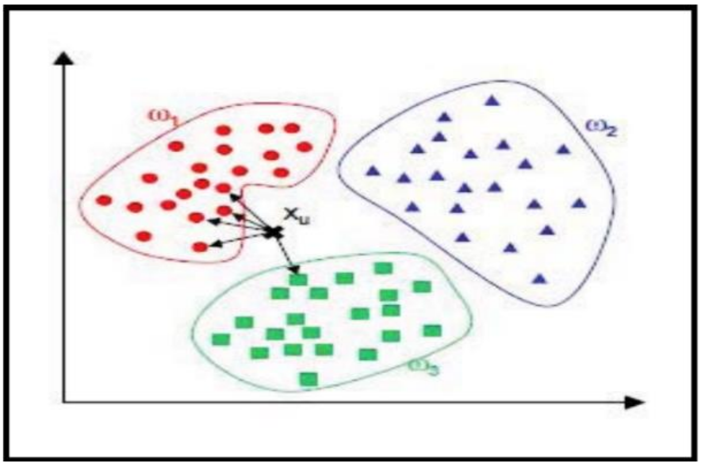

Figure 22 Example of K-Nearest Neighbour

3.3.5 Self-Organizing maps (SOM)

An important ability of neural networks (NN) is of error forbearing [28]. Comparable to brain a NN haven't get posh by minor irregularities. Due to its rapid learning capability it alter itself competently with respect to the data. Over-all, SOM is a sort of neural based network that uses a kind of unsupervised learning technique. It is called Map as it tries to configure its coefficients to track given input data. The SOM nodes try to develop themselves like the inputs. Lesser the difference more the SOM is learnt. Similar to any other neural network SOM also reduces the dimensionality of data as well as it reduce the overall complexity.

3.3.5.1 Organization of a self-organizing map

The SOM's arrangement is very simple, can be imagined with the help of Figure 23 where a SOM network of size 4x4 is depicted. Every node is connected to each input whereas there is no connection among the nodes. Each node can refer to a distinct format (i,j). SOM node is the fundamental part of a body. Each node contains a set of weights that is equal to the input vector weight.

3.3.5.2 SOM Algorithm

There are mainly 6 steps of SOM algorithm [29]:

1. Initialize each node with a random weight.
2. A vector is given as input to the network as a training data.
3. Every node is scanned to compute change with respect to the input vector. The winner is of least distance. The change is calculated by the following formula.

$$Dist\ from\ Input^2 = \sum_{i=0}^{n} (I_i - W_i)^2$$

n = number of nodes

I = current input vector

W = node's weight vector

4. The area of locality of the least distanced node is computed. Initialized with the radius and contracts on every repetition. Radius of neighboring node is calculated by

$$\sigma(t) = \sigma_o e^{-\frac{t}{\lambda}}$$

t = current iteration

λ = time constant = numIterations / mapRadius

σ_o = radius of the map

5. Nodes within a radius different to the input vectors are adjusted. A node that is nearer with respect to the winner, the more its coefficients are changed. New weight is evaluated using equation.

$$W(t+1) = W(t) + \Theta(t)L(t)\big(I(t) - W(t)\big)$$

Learning Rate is calculated using $L(t) = L_0 e^{-\frac{t}{\lambda}}$

6. Step (2) to (6) is executed for N repetitions.

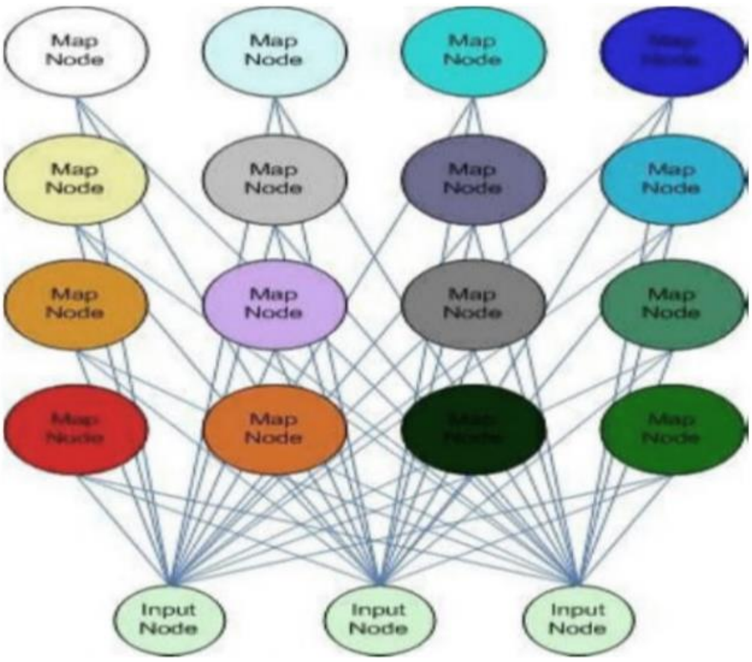

Figure 23 Structure of a SOM [29].

3.3.6 Deep Learning

Deep learning is inspired mostly by neurons and brain and based on very lose simulation of brain. The algorithms uses several levels and layers of raw data to learn automatically by using a deep structure of neural networks make up of many hidden layers. The good thing about deep learning algorithm is that it automatically mine features that are more related to classification and involves meaningful information which is not depend upon other features. The DBN model is based on a number of Restricted Boltzmann Machine (RBM) which uses unsupervised learning technique [30].

Every single hidden layer in DBN has an RBM with no connection with the units of the same layer and have indirect and even links to the units available in the visible layer in order to make the calculation of conditional probabilities easy. RBM is trained and weights are assigned to each units and the main issue is weights are not generative as shown in Figure 24. In Figure 24. w_{nm} Represents the weights between the hidden layer h_n and visible layer v_m [31]. The RBM type used in this work is Gaussian and its structure is shown in Figure 24:

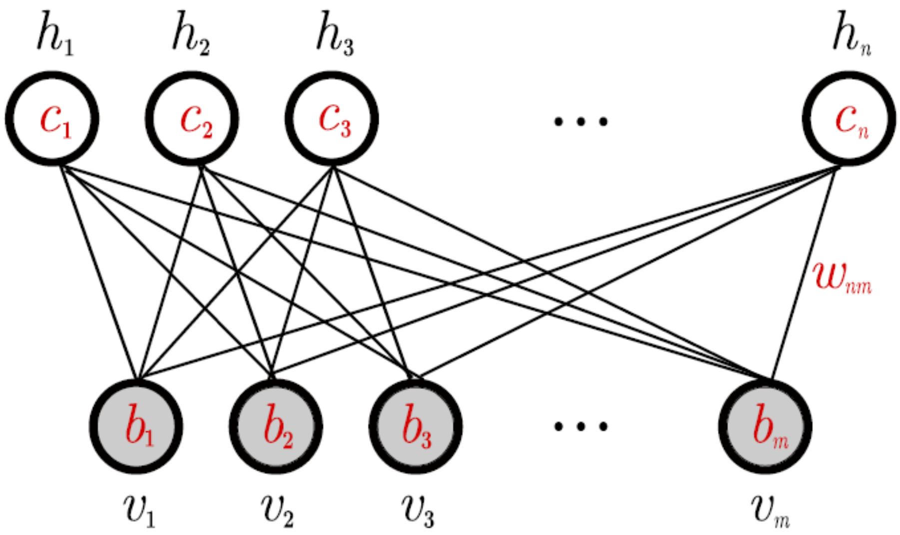

Figure 24 Organization of RBM with n hidden and m visible layers

3.3.6.1 Deep Belief Network

Consider an RBM based on of visible units v and hidden units h in Deep Belief Network (DBN). The joint probability distribution is given by [31]:

$$p(v) = \frac{\sum e^{-E(v,h)}}{\sum_u \sum_g e^{-E(u,g)}}$$

Where E is given by

$$E(v,h) = \sum_{j \in Visible} a_j v_j - \sum_{j \in hidden} b_j h_j - \sum_{i,j} v_j h_j \omega_{ij}$$

Where v_j, h_j are the states in binary for hidden unit j and visible unit i. $\omega_{ij}$is the weight between the hidden unit j and visible unit i and $a_j$and $b_j$are their biases. E is the energy function based on visible and hidden units. After computing the energy function the next step is to assign probability to a pair of hidden unit and visible unit based on energy function:

$$p(v,h) = \frac{1}{z} e^{-E(v,h)}$$

Once the probabilities are assigned to the network based on training data it can be optimize by tuning the weights and lower the energy of biases. The rate of change of logarithmic probabilities of training with previously calculated weights is:

$$\frac{\partial \log p(v)}{\partial \omega_{ij}} = \frac{\sum_{v\epsilon\, D} \partial \log p(v)}{\partial\ \omega_{ij}} = E_{data}\left[\frac{\partial E(v,h)}{\partial \omega_{ij}}\right] - E_{model}\left[\frac{\partial E(u,g)}{\partial \omega_{ij}}\right]$$

Where E is the expected value respond to training set D and the conditional probability distribution of dataset on *p(h|v)* is used to sample the hidden variables.

$$p(h_j = 1|v) = \sigma(b_j + \sum v_i \omega_{ij})$$

Where $\sigma(x) = \frac{1}{1+\exp(-x)}$ is the logistic sigmoid function, for training and RBM classifier the energy function is given by:

$$E(v,l,h) = -\sum_i \alpha_i v_i - \sum_j b_j h_j - \sum_{i,j} \omega_{ij} - \sum_y c_y l_y - \sum_{y,j} \omega_{yj} h_j l_y$$

Where l is the binary class label, the visible unit vector is concatenated with l class labels.

3.3.6.2 Boosting Deep Belief Network Classifier

The concept of boosting has been used to make weak classifier powerful. We have used the idea of Ada-boost algorithm to boost weak classifier. The EEG channel selected for analysis are C3, C4 and Cz and boosting has been performed as [32] [33]:

$$M_k(X) = \sum_{k=1}^{k} c_k f_k(X)$$

Where c_k is the estimated coefficient weights for each DBN model and for each input data DBN model produces discrete classification [33].

3.4 Evaluation

To find out how good a classifier is? Different techniques has been used to measure the performance of a classification algorithm. In this thesis, we used classification accuracy, Mini batch mean square error and full batch mean square error to evaluate classifier performance. The explanation about these methods are given in the following section.

3.4.1 Classification Accuracy

The mostly used and an easy way to rate the classifier is to use classification accuracy (Acc) as a method to evaluate the EEG signal classification. It is calculated as follows.

$$Acc = \frac{\sum_{i=1}^{M} n_i}{N}$$

There are some limitation related to classification accuracy. Firstly it does not consider off-diagonal results which is consider in confusion matrix. Secondly it has been seen that its value depends upon number of samples in the class [34].

3.4.2 Mean Square Error (MSE)

The average of the square of the difference of estimated value and original value is known as the mean squared error (MSE).If $\hat{Y}$ is a vector set of n estimates, and Y is the vector set of the correct values, then the (expected) MSE of the predictor is:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y_i} - Y_i)^2.$$

We have used mean square error for full batch and mini batch evaluation and record their results. In full batch whole dataset has been used for evaluation and in mini batch a chunk of data has been given to measure error results.

Chapter 4

EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experiment and Dataset Details

In this section, we present the experimental results and details related to the experiment.

4.1.1 Dataset Details

The dataset named Graz data from BCI competition 2003 has been used in classification for training and testing purposes. The dataset was collected from a subject comforting on a chair with support to his arms. The objective is to move a block in the vicinity of EEG signals comprising of left and right hand movement. The electrodes are placed on scalp illustrated in Figure.25.

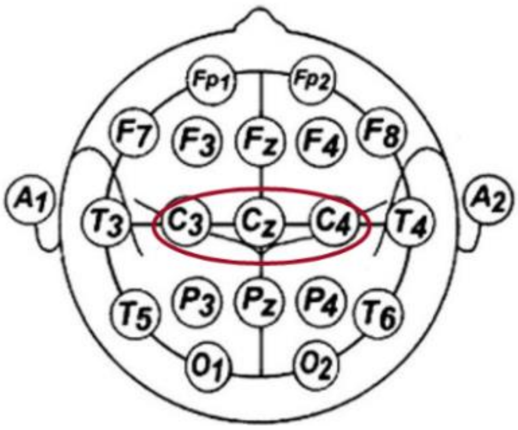

Figure 25 Electrode placement based on the experiment

The database contain 280 trails out of which 140 correspond to training set and rest of them correspond to testing signals. Each trail last for 9 seconds containing data of Cz, C4and C3. The movement of experimental stimulus is shown in Figure 26. The sampling rate is 128Hz. Low frequency brain signals lie in the range of 0.3-40Hz. Therefore a frequency range of 25Hz i.e. 0.5-30 Hz is extracted through a band-pass filter [10].

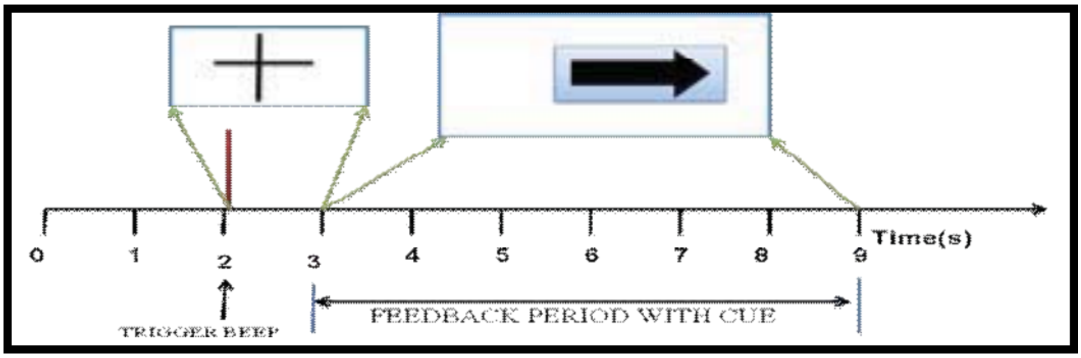

Figure 26 Visual stimuli along with timing scheme

4.2 Experimental Flow

The flow is divided in to two categories. One is for Self–Organizing Maps and the other one is for Deep Learning.

4.2.1 Self-Organizing Maps based neural Network

The input vector of SOM is the feature vector. A total of 140 feature vectors has been given to the network for training purpose. A suitable size SOM has been selected. A vector of Weight Array has been constructed with respect to the dimension of SOM network with a length same as of input vector. All vectors are generated randomly according to the weight or coefficients array [8]. The network has trained for all input vectors for large repetition and after that the error has been computed, known as Average Error. The whole process for SOM has been shown in Figure 27.

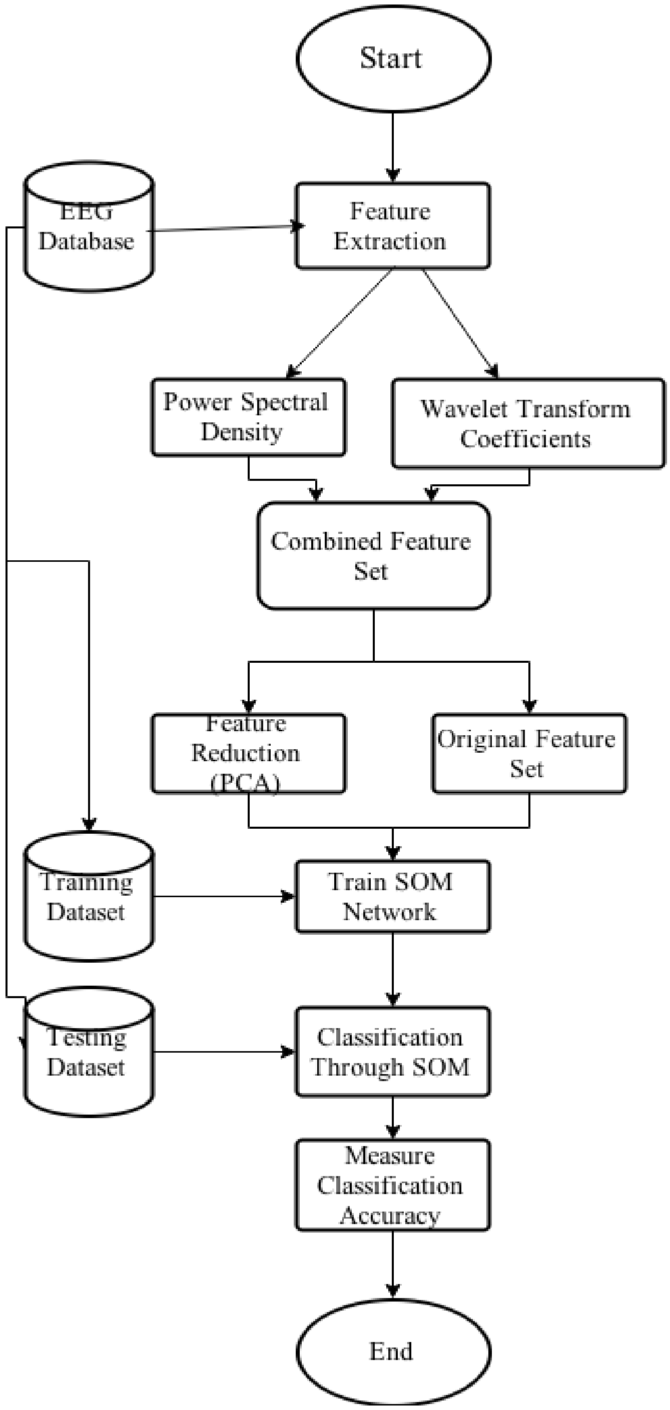

Figure 27 Flow Chart of SOM based Classification

4.2.1.1 Feature vector set

The data taken for features extraction is from t = 3s to 9s. The signal has a frequency range 0.5-30Hz. The feature vector consist of wavelet coefficients, PSD estimates for both bands i.e. (8-12Hz and 18-25Hz) and their corresponding powers. These steps has been performed in MATLAB using the toolbox of wavelet and signal processing (Table 2).

Table 2 Feature Sets With Size

Features	Dimension (Features × samples)
Bior6.8 Wavelet Coefficient	102 ×140
PSD estimate	768 ×140
Mean Power of signal	1 ×140
Total features	871 ×140

The data size 871x 140 has been termed as non-reduce feature set. After applying Principal Component Analysis (PCA) except the mean power, the reduced feature vectors came of size 91×140. Both the feature sets have been given to different classifier for training and testing purposes.

4.2.1.2 Performance Analysis

The both features vectors has been provided to the above mention classification algorithms using MATLAB. The classification results of both reduced and non-reduced feature vectors has been shown in Table 3.

Table 3 Result of Classification

No.	Classification Algorithms	Accuracy (original) %	Accuracy (reduced) %
1	LDA	80.30	82.64
2	QDA	80.50	81.70
3	KNN	77.50	82.90
4	SVM (Linear)	81.42	81.42
5	SOM	83.45	84.17

It can be seen from Table II that SOM based approach perform quiet good in both cases, SOM based neural network classifier gives maximum classification results of 83.45%. However, there has been a raise in performance accuracy compared to [10] by simply changing the wavelets type from Daubechies to bior6.8, also, kNN has displayed noteworthy rise in the classification results from 77.50% to 82.90%.

4.2.2 Deep Belief Nets (DBN)

The experiment has been carried out on three different feature set and results for all of the cases has been compared and discussed. The algorithm flow is shown in Figure 28.

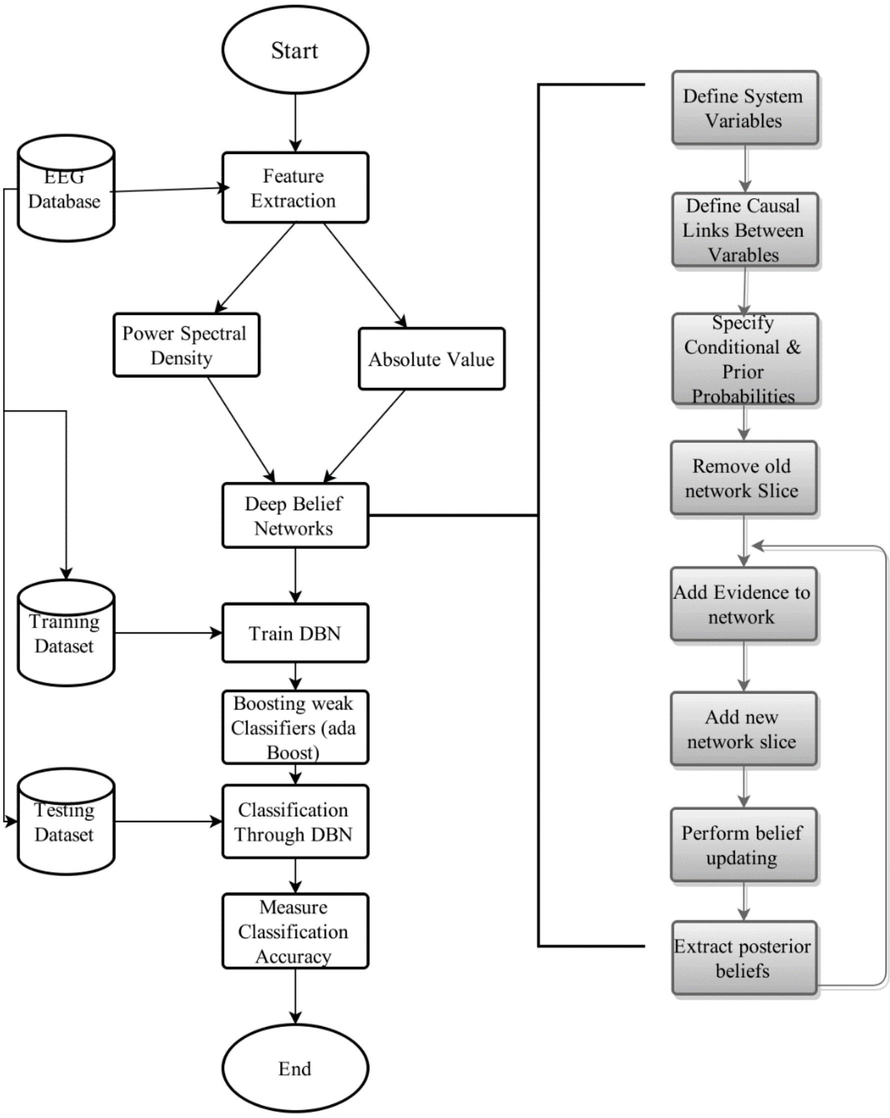

Figure 28 Flow Chart of DBN based Classification

The first part is of feature extraction from the dataset. Once the desired features has been extracted from the dataset i.e. Power Spectral Density and Aggregated EEG signal, the next step is to train a deep belief network with this feature set. The training of Deep Belief Network has been carried out on the training dataset. The DBN has been trained for both PSD and Aggregated EEG signal separately and their outputs has been compared. After the training of DBN the classifier has been boosted with ada-boost algorithm to make weak classifiers strong. The last step is to classify the test dataset using DBN and results has been compared.

4.2.2.1 Experimental Parameters

From 140 total samples 70 samples has been used for training and 70 for testing purpose. Each node has been initialized with a random weight.

The turning parameters were set as:

Learning rate for weight and biases = 0.07

Momentum=0.5

Weight decay = 0.002.

Range of hidden layers for DBN has been set from 4 to 20 for training purpose and has been tested for every sample and result for every layer has been enlisted in table 1-4. The unit size is fixed for every layer i.e.10 unit in each layer.

4.2.2.2 Experimental Results

The tables given below contain results of Mean square error for EEG signal classification through DBN. Three different features has been used to validate the DBN. The first feature set is dependent on time domain and the other two features set are from the frequency domain.

The Table 4 and Table 5 shows the results of time domain features in which the aggregation of 3 EEG channels has been carried out and then it has been given to DBN for training and Classification purpose. Two different types of validation techniques has been used to check the results i.e. Mini Batch and Full Batch Mean Square Classification Error. The other two features depend upon the frequency domain, table 3 show the classification error for Power Spectral Density (PSD) and table 4 shows the results for Fourier Transform (FFT).

Table 4 shows the results of mini batch means square classification error when DBN is trained with aggregated EEG signals. The results shows that as we propagate along the hidden units, relatively odd number of hidden units gave less classification error as compare to even number of hidden layers. As we move along the epoch it is clearly shown that with every epoch the error reduced and DBN learning is improving with each epoch.

Table 4 Mini Batch Mean Square error using aggregated absolute value

Hidden layers	Epoch						
	1	4	6	7	8	10	Error
4	0.32365	0.22715	0.19225	0.10380	0.05547	0.02137	7%
6	0.49859	0.49618	0.40745	0.26631	0.25521	0.26630	29%
7	0.30189	0.29237	0.20115	0.10559	0.05974	0.01370	8%
9	0.36111	0.2511	0.24729	0.17114	0.14391	0.05030	11%
10	0.38963	0.37796	0.23674	0.21644	0.18168	0.08136	16%
12	0.37125	0.25995	0.19381	0.18716	0.13798	0.07046	11%
14	0.39381	0.27904	0.24590	0.23087	0.21993	0.14084	17%
15	0.35488	0.22000	0.11820	0.11339	0.07837	0.02874	6%
17	0.30835	0.25203	0.21098	0.19075	0.17331	0.07226	11%
18	0.28049	0.25334	0.25358	0.25196	0.23529	0.16987	14%
19	0.27582	0.24876	0.15606	0.14082	0.10631	0.03305	8%
20	0.31849	0.26069	0.24548	0.24479	0.25184	0.21088	16%
Epoch wise	30%	23%	17%	13%	10%	4%	7%

Full batch mean square classification error has been recorded in Table 5 with aggregated signal as feature set. The numbers shows a relatively less error compare to mini batch classification error with a minimum error reading of 4% for a DBN of 15 hidden layers. The results shows that after each epoch the error reduces with a minimum error of 2% in epoch 10 for any hidden layers and the overall average of full batch mean square error for all cases is recorded 5%.

Table 5 Full Batch Mean square Error using Aggregated Absolute Value with different number of hidden layers

Hidden layers	Epoch						
	1	4	6	7	8	10	Error
4	0.328874	0.186094	0.127025	0.055325	0.032907	0.010847	6%
5	0.303948	0.250407	0.211553	0.251637	0.123172	0.040178	11%
6	0.498398	0.494803	0.264515	0.256402	0.254118	0.222160	28%
7	0.260469	0.229944	0.100652	0.055693	0.041782	0.012354	6%
9	0.279942	0.233898	0.164947	0.198849	0.100792	0.025755	8%
10	0.375353	0.353728	0.192056	0.174571	0.112999	0.050374	14%
12	0.249831	0.244736	0.175301	0.140486	0.101412	0.045612	8%
14	0.399336	0.280878	0.238957	0.242188	0.224081	0.114004	15%

15	0.238821	0.204236	0.087956	0.130813	0.035063	0.035500	4%
16	0.412291	0.238595	0.162861	0.115279	0.067042	0.023590	7%
18	0.249528	0.253878	0.240352	0.221536	0.210868	0.127481	12%
19	0.285597	0.206465	0.125359	0.103638	0.066685	0.019502	5%
20	0.253958	0.273584	0.243070	0.248365	0.228325	0.184926	15%
Epoch wise	27%	21%	13%	12%	7%	2%	5%

The DBN classification results were relatively bad when Power Spectral Density has been used as a feature set for DBN. The full batch classification error using power spectral density has been recorded in Table 6. The result shows a minimum classification error of 15% for the hidden layer 13. The classification error reduces after each epoch but the reduction rate is less compare to Table 4 and Table 5. The overall average classification error is of 18 % has been recorded for the PSD feature set.

Table 6 Full Batch Classification Error Using PSD

Hidden Layers	Epoch						
	1	3	5	6	8	10	Error
4	0.158124	0.197538	0.171107	0.151335	0.150454	0.155412	17%
5	0.276637	0.161026	0.152534	0.166176	0.18365	0.149961	18%
8	0.151637	0.157858	0.16407	0.159322	0.194539	0.151622	17%
9	0.299362	0.275149	0.286839	0.152936	0.151099	0.161565	23%
10	0.156181	0.150529	0.153986	0.150587	0.15969	0.153361	16%
13	0.150986	0.150162	0.156962	0.178769	0.151577	0.150494	15%
14	0.297748	0.189421	0.150261	0.169156	0.150458	0.153936	18%
15	0.396984	0.165612	0.158541	0.183391	0.156187	0.15078	21%
16	0.301985	0.229711	0.168155	0.155346	0.162087	0.155824	19%
17	0.240276	0.201087	0.173229	0.161217	0.204058	0.159594	18%
18	0.386699	0.167499	0.162614	0.153515	0.179621	0.159482	20%
20	0.239277	0.161283	0.160649	0.156961	0.156937	0.217208	18%
Epoch Wise	25%	18%	18%	16%	17%	16%	18%

It has been depicted from Table 4 to 6 that DBN results are very worthy when the feature set is in the time domain. The features that are related to frequency domain doesn’t perform well i.e. the minimum mean square classification error for time domain based aggregated feature set is

7% in mini batch and 5% in full batch, but for frequency domain feature set the minimum error is 15 % for Power Spectral Density.

In Figure 30, comparison of mini batch and full batch classification error has been shown through which we conclude that the difference between these two validation methods is quite low so the DBN based classification is equally good for both validation methods.

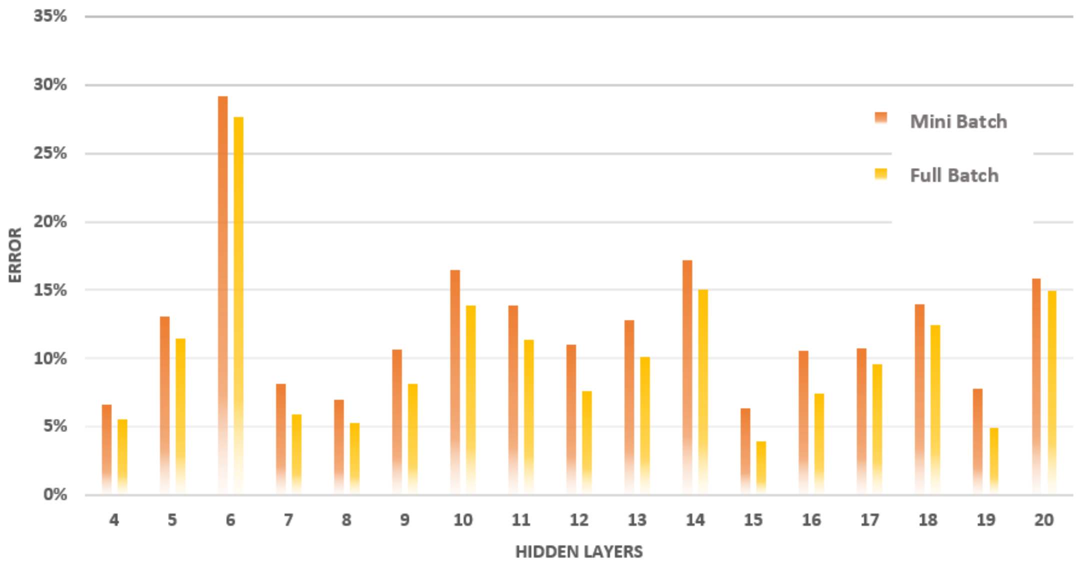

Figure 29 Comparison of mini and full batch classification error Based on Hidden layers

The comparison of PSD and Aggregated Feature set has been shown in Figure 31. It has been depicted for graph that the time domain related features work well for DBN as compare to frequency domain feature like PSD and FFT. The DBN with hidden layer 8 and 12 perform well in both cases. The minimum error for aggregated signal feature set is 4% for the 15 hidden layers and for PSD it is minimum for 10 hidden layer.

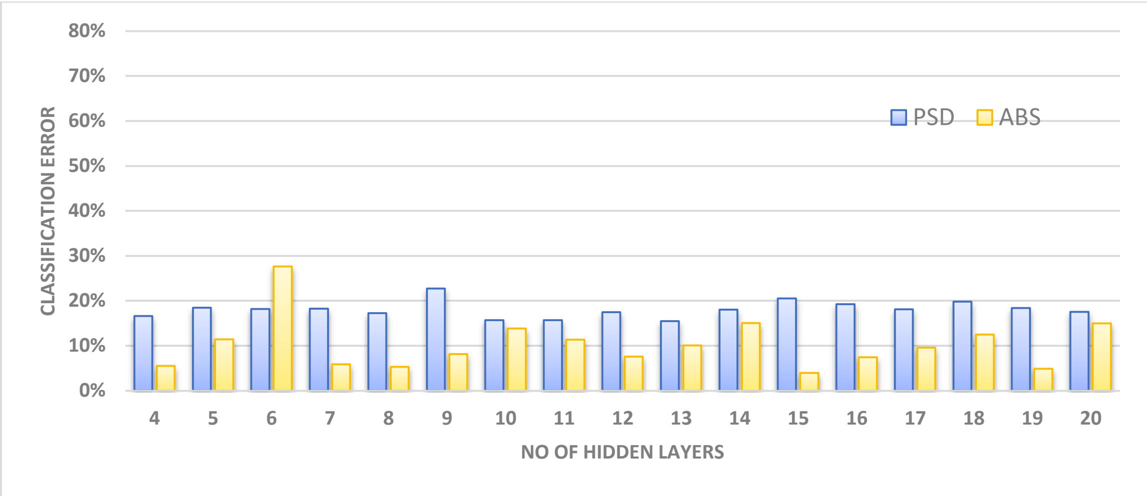

Figure 30 Classification error of three different features set

Chapter 5

CONCLUSION

5.1 Conclusion

In this thesis, state of the art classification techniques have been compared using motor imagery based EEG signals. The overall process can be summarized as follow.

- For signal enhancement: for extracting required signal band-pass filtering has been used, Median filtering has been used to remove pulse noise.
- For extracting features: Wavelet Transforms (WT), Power Spectral Density (PSD), Aggregated EEG Signal and Principal Component Analysis (PCA).
- For classification: Support Vector Machines (SVM), Deep Belief Nets (DBN), Self-Organizing Maps (SOM)
- For evaluation: classification accuracy, Mini Batch Mean Square Error, Full Batch Mean Square Error.

All these methods are experimented on BCI Competition 2 dataset and results have been recorded. We present an efficient approach to classify motor imagery EEG signals with supervised and unsupervised learning algorithm by extraction features that found to be the best features for classification.

For SOM based classification techniques, the feature set includes Bior 6.8 Wavelet transform, PSD approximation and mean power. A comprehensive analysis has been presented and it has been concluded that SOM gave the highest classification efficiency compared to earlier discussed algorithms [35, 36] which is also authenticated in many writings [37, 38 and 39]. In most of the cases, the classification of reduced feature set by PCA has increased as compare to the non-reduced feature sets, which concurrently enhances the classification accurateness. It has also been evident from the results that by changing wavelet transform from Daubechies of order 4 to bi-orthogonal wavelets the accuracy has been increased almost 2%.

In the second part, Deep Belief Network classification model for the recognition of Motor Imagery based EEG pattern is proposed and tested. The experimental results in Figure 32 displayed a consistent enhancements for all verified cases over different state of the art algorithms like LDA, QDA, kNN, SVM, Naïve and other algorithms given in article [36] and also outnumber the classification results for [40] also through multiple cross-validation experiments.

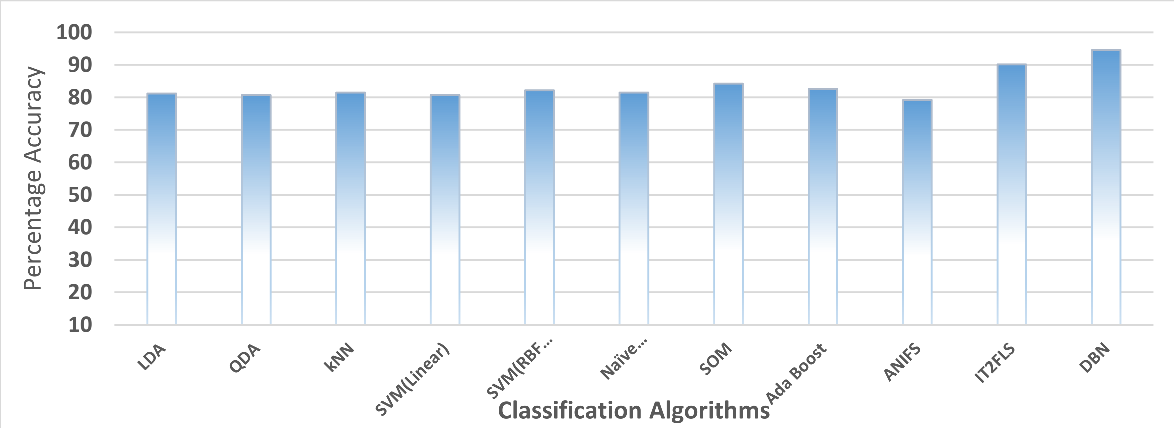

Figure 31 Comparison of different classification Algorithms

The test has been carried out on different number of hidden units, and it has been concluded that the number of nodes had no significant influence on the performance of classification. The detailed experimental results have shown that Deep Learning algorithm and more specifically Deep belief networks can also perform very efficiently in classifying the tasks related to Motor Imagery based EEG signals. The results of the experiment also shows that time domain features that have been aggregated along all three channels have a great concentration on the result accuracy. So it can easily be learned from the experiment shown above that deep learning can play a vital role in classifying different processes related to EEG signals with a capability to train huge dataset timely and easily and it can also serve as a powerful classification tool for BCI development.

5.2 Future work

The proposed Deep Learning approach is relatively new, robust and adaptive as compare to the other reported approaches, so in order to drive EEG sourced BCI devices (mobile robot) it is a very good approach which require less computation and give maximum efficiency.

Our future plan is to implement this classification technique on a real system and design a system that has the ability to online classify motor imagery EEG signals and able to control a mobile robot in a real environment. Furthermore, we are planning to implement morphological signal processing to enhance EEG signal and develop an algorithm that has the ability to adaptively select the features and combined the result of more than one classification algorithm to decide the required action.

Publications

Journal Publications:

1. Baig, Muhammad Zeeshan, Yasar Ayaz, Waseem Afzal, Syed Omer Gillani, Muhammad Naveed, and Mohsin Jamil. “A Comparative Analysis of Motor Imagery based EEG Signals Classification Algorithms with Deep Belief Networks.” In BioMed Research International Journal (Submitted) (ISI Indexed)
2. Baig, Muhammad Zeeshan, Yasar Ayaz, Waseem Afzal, Syed Omer Gillani, Muhammad Naveed, and Mohsin Jamil. “Motor Imagery Based EEG Signal Classification Using Self Organizing Maps.” In Science International Journal (Published) (ISI Indexed)

Conference Publications:

1. Baig, Muhammad Zeeshan, Ehtasham Javed, Yasar Ayaz, Waseem Afzal, Syed Omer Gillani, Muhammad Naveed, and Mohsin Jamil. "Classification of left/right hand movement from EEG signal by intelligent algorithms." In Computer Applications and Industrial Electronics (ISCAIE), 2014 IEEE Symposium on, pp. 163-168. IEEE, 2014. (Published)

Bibliography

[1] Sitaram, Ranganatha, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler, and Niels Birbaumer. "FMRI brain-computer interface: a tool for neuroscientific research and treatment." Computational intelligence and neuroscience 2007 (2007).

[2] Coyle, Shirley M., Tomás E. Ward, and Charles M. Markham. "Brain–computer interface using a simplified functional near-infrared spectroscopy system." Journal of neural engineering 4, no. 3 (2007): 219.

[3] Hill, N. Jeremy, Thomas Navin Lal, Michael Schröder, Thilo Hinterberger, Guido Widman, Christian E. Elger, Bernhard Schölkopf, and Niels Birbaumer. "Classifying event-related desynchronization in EEG, ECoG and MEG signals." In Pattern Recognition, pp. 404-413. Springer Berlin Heidelberg, 2006.

[4] Pineda, Jaime A. "The functional significance of mu rhythms: translating “seeing” and “hearing” into “doing”." Brain Research Reviews 50, no. 1 (2005): 57-68.

[5] Daly, Janis J., and Jonathan R. Wolpaw. "Brain–computer interfaces in neurological rehabilitation." The Lancet Neurology 7, no. 11 (2008): 1032-1043.

[6] Mak, Joseph N., and Jonathan R. Wolpaw. "Clinical applications of brain-computer interfaces: current state and future prospects." Biomedical Engineering, IEEE Reviews in 2 (2009): 187-199.

[7] van Gerven, Marcel, Jason Farquhar, Rebecca Schaefer, Rutger Vlek, Jeroen Geuze, Anton Nijholt, Nick Ramsey et al. "The brain–computer interface cycle." Journal of Neural Engineering 6, no. 4 (2009): 041001.

[8] Niedermeyer, Ernst, and FH Lopes da Silva, eds. Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins, 2005.

[9] “Electroencephalography”, http://en.wikipedia.org/wiki/Electroencephalography, last visited on 15/04/2015.

[10] “Intendix,” http://www.intendix.com/, last visited on 15/04/2015.

[11] Decety, J., & Ingvar, D. H. (1990). Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation. Acta Psychologica, 73, 13-24.

[12] Decety, J., & Stevens, J. (2009). Action representation and its role in social interaction. In K.D. Markman, W.M.P. Klein & J.A. Suhr (Eds.), The Handbook of Imagination and Mental Simulation. New York: Psychology Press.

[13] Brumberg, Jonathan S., Alfonso Nieto-Castanon, Philip R. Kennedy, and Frank H. Guenther. "Brain–computer interfaces for speech communication." Speech communication 52, no. 4 (2010): 367-379.

[14] "P3SpellerTask," http://www.bci2000.org/wiki/index.php/UserReference:P300ClassifierMethods , last visited on 15/04/2015.

[15] Cincotti, Febo, Donatella Mattia, Fabio Aloise, Simona Bufalari, Gerwin Schalk, Giuseppe Oriolo, Andrea Cherubini, Maria Grazia Marciani, and Fabio Babiloni. "Non-invasive brain–computer interface system: towards its application as assistive technology." Brain research bulletin 75, no. 6 (2008): 796-803.

[16] Pires, Gabriel, and Urbano Nunes. "A Brain Computer Interface methodology based on a visual P300 paradigm." In Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on, pp. 4193-4198. IEEE, 2009.

[17] Palankar, Mayur, Kathryn J. De Laurentis, Redwan Alqasemi, Eduardo Veras, Rajiv Dubey, Yael Arbel, and Emanuel Donchin. "Control of a 9-DoF wheelchair-mounted robotic arm system using a P300 brain computer interface: Initial experiments." In Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on, pp. 348-353. IEEE, 2009.

[18] Xu, Huaiyu, et al. "Feature extraction and classification of EEG for imaging left-right hands movement." Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on. IEEE, 2009.

[19] Bhattacharyya, Saugat, Anwesha Khasnobish, Amit Konar, D. N. Tibarewala, and Atulya K. Nagar. "Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms." In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on, pp. 1-8. IEEE, 2011.

[20] Mu, Zhendong, Dan Xiao, and Jianfeng Hu. "Classification of motor imagery EEG signals based on STFTs." Image and Signal Processing, 2009. CISP'09. 2nd International Congress on. IEEE, 2009.

[21] Herman P., Prasad G., McGinnity T.M., Coyle D. “Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification.” IEEE Trans. Neural sys. Rehab eng. 16(4), August 2008, pp. 317-326

[22] Pearson, Karl. "LIII. On lines and planes of closest fit to systems of points in space." The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2, no. 11 (1901): 559-572.

[23] Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. John Wiley & Sons, 2012.

[24] Srivastava Santosh, Gupta Maya R., Frigyik B´ela A., “Bayesian Quadratic Discriminant Analysis” Journal of Machine Learning Research 8 (2007) 1277-1305

[25] Baoli L., Shiwen Y., Qin L. “An improved K- nearest neighbor algorithm for text categorization.” Proceedings of the 20th International conference on computer processing of Oriental languages,Sheyang, China, 2003.

[26] Jakkula Vikramaditya, “Tutorial on Support Vector Machine (SVM)”.

[27] Gunn Steve R.,”Support Vector Machines For Classification and Regression”, Technical report, University of Southampton, May 1998.

[28] James L. Noyes.Artificial Intelligence with Common Lisp: Fundamentals of Symbolic and Numeric Processing. D.C. Heath, Lexington, MA; 1992.

[29] Ul Hassan, Shabih, et al. "Modified SOM based intelligent semi-autonomous navigation system." Neural Network Applications in Electrical Engineering (NEUREL), 2012 11th Symposium on. IEEE, 2012.

[30] Yoshua, B., Pascal, L.: Greedy layer-wise training of deep networks. NIPS (2006)

[31] Gorge, H.S.O.: A fast learning algorithm for deep belief nets. Neural Computation, 1527–1554 (2006)

[32] Kirkup, L., Searle, A.: EEG-based system for rapid on-off switching without prior learning. Medical and Biological Engineering and Computing, 504–509 (2007)

[33] Hochberg, L., Serruya, M.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 164–170 (2006)

[34] Dornhege, G., J. del R Millán, T. Hinterberger, D. McFarland, and K. Müller. "Evaluation Criteria for BCI Research." 327-342.

[35] Bhattacharyya S., Khasnobish A., Konar A., Tibarewala D.N., “Performance analyisis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data”, International conference on systems in medicine and biology, IIT Kharagpur, 2010

[36] Khasnobish A., Bhattacharyya S., Konar A., Tibarewala D.N., “K- Nearest neighbor classification of left-right limb movement using EEG data.”, International conference on Biomedical Engineering and assistive technologies, NIT Jalandhar, 2010.

[37] Thomas K.P., Guan C., Tong L.C., Prasad V.A. “An adaptive filter bank for motor imagery based brain computer interface.” 30th Annual International IEEE EMBS Conference, Vancouver, Canada, August, 2008

[38] Wang Y., Makeig S. “Predicting Intended Movement Direction using EEG from Human Posterior Parietal Cortex.” HCI international, San Diego CA., July 2009.

[39] Ince N.F., Goksu F., Tewfik A.H., Arica S., “Adapting subject specific motor imagery EEG patterns in space-time-frequency for a brain computer interface.” Biomedical signal processing and control 4, 2009, 236-246

[40] Nguyen, T., Khosravi, A., Creighton, D., & Nahavandi, S. (2015). EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems. Expert Systems with Applications, 42(9), 4370-4380.