MOTOR IMAGERY BASED EEG SIGNAL CLASSIFICATION USING SELF ORGANIZING MAPS

***Muhammad Zeeshan Baig, Yasar Ayaz, Syed Omer Gillani, Mohsin Jamil and Muhammad Naveed**
Robotics and Intelligent System Engineering (RISE) Lab
School of Mechnical and Manufacturing Engineering (SMME)
National University of Science and Technology H-12 Islamabad, Pakistan
*Contact: zeeshan@jinnah.edu.pk

ABSTRACT: Classification of Motor Imagery (MI) tasks based EEG signals effectively is the main hurdle in order to develop online Brain Computer interface (BCI). In this research article, 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 bi-orthogonal Wavelet Transform (WT) coefficients, Welench Power Spectral Density estimates and the average power. In this study, we have presented a comparison of mostly used classification algorithm with a new unsupervised learning technique for classification i.e. Self-organizing maps (SOM) based neural network. SOM and other algorithms have been used to categorize the feature vector acquired from the EEG data-set; into their corresponding classes. Both orignal and reduced feature set has been used for classification of motor imagery based EEG signals. The reduction is performed by applying Principal Component Analysis (PCA). It has been depicted from measured data that SOM shows a maximum classification accuracy of 84.17% on PCA implemented reduce feature set. Furthermore, an 2% increase in classification accuracy has been attained by using bi-orthogonal wavelet transform instead of Daubechies WT.

Keywords: EEG, SOM, Motor Imagery.

INTRODUCTION

The Brain Computer Interface (BCI) is a device that permit brain signals to interact with the environment. BCI has been divided into two groups namely, invasive and non-invasive [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.

There are many sources to measure brain activities for BCI i.e. EEG, ECoG, fMRI, MEG, LPF [3]. The BCI system with EEG input has been the most reliable and frequently used to measure brain activity due to the non-invasive EEG electrodes availability. It also exhibit high temporal resolution.

Several channel electrodes (14, 64, and 128 etc.) are available in market that can be used to acquire EEG signals. Authors have suggested to use C3, C4 and Cz to control the motor imagery related BCI.

The main step after signal acquisition is to extract dominant features. The most widely used features are mean, variance, short time Fourier transforms (STFT), standard deviation, Recursive energy efficiency (REE), wavelet transform (WT) and Hjorth parameters [5-6]. Once the features are extracted next big hurdle is to classify these features efficiently with maximum accuracy in order to make an online BCI. The features vector dimensions can be reduced by applying PCA or ICA.

BCI performance is measured by its classification accuracy. In order to make online classification the classifiers must be quick enough to do real time classification of the EEG signals. Mostly used classifiers are k-nearest neighbour (KNN) or support vector machine (SVM) [6]. The main objective of this writing is to show a new classification approach i.e. Self-organizing map based neural network 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.

The related work regarding self-organizing maps is elucidated in section II. The pre-processing containing the experimental setup use to collect the database, feature extraction concisely defined in section III. Feature classification algorithms are defined in section IV. Section V is based on the result of different classifier that has been used. Conclusions are enumerated in section VI.

Related work

Self-Organizing maps (SOM)

An important ability of neural networks (NN) is of error forbearing [7]. Comparable to brain a NN haven't got posh by minor irregularities. Due to its rapid learning capability it alters 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 reduces the overall complexity.

Organization of a self-organizing map

The SOM's arrangement is very simple, can be imagined with the help of Fig 1 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.

SOM Algorithm

There are mainly 6 steps of SOM algorithm [8]:

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.
he 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
$\sigma_o \sigma_o$ = 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 $L(t) = L_0 e^{-\frac{t}{\lambda}}$

$$L(t) = L_0 e^{-\frac{t}{\lambda}}$$

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

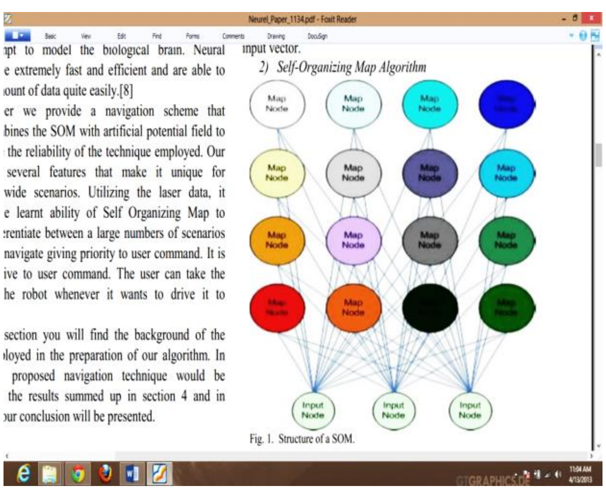

Figure 1 Structure of a SOM [8].

Pre-processing

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 Fig.2.

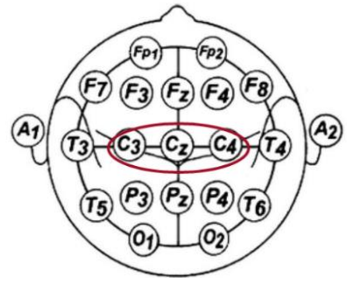

Figure 2 Electrode placement based on the experiment

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 Fig.3. 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].

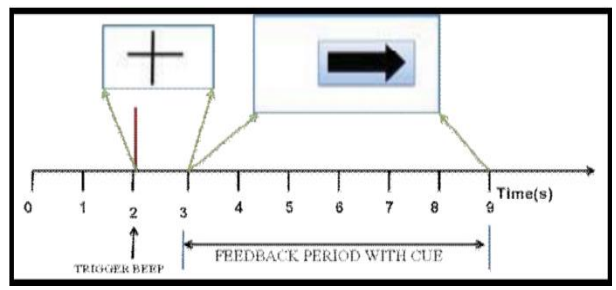

Figure 3 Visual Stimuli along with Timing Scheme

Feature vectors extraction

Feature vectors have been extracted from the predefined channels C3 and C4 [9]. The feature vector based on WT and statistical parameters of the selected EEG channels has been used by saugat in [10] with a little modification in using wavelet transform. We have used the same features In order to compare the predefine techniques.

Wavelet transform

The inability to tackle non-stationary signals has been the main hurdles in Fourier transform (FT) as it neglects the small changes in high frequency components [4]. 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 [10] to acquire the frequency bands signals.

The wavelet function $\psi(t) \epsilon L^2(R)$ $\psi(t)\epsilon L^2(R$ $\psi(t)\epsilon L^2(R$ 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 Fig. 4 and Fig. 5.

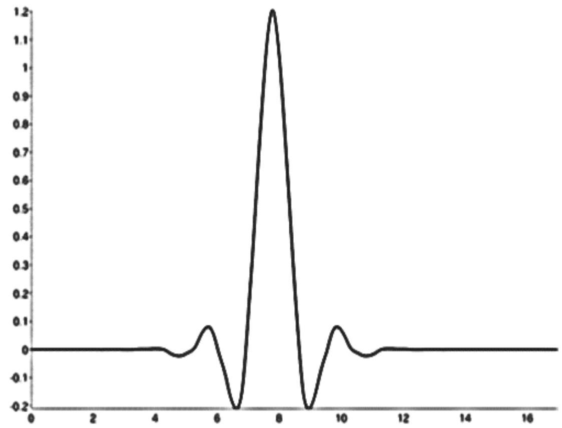

Figure 4 Decomposition scaling function φ

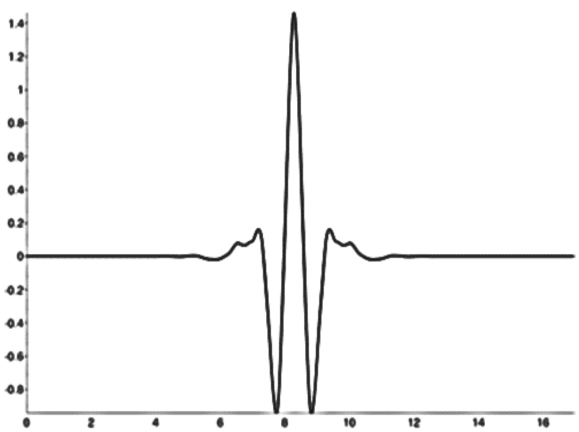

Figure 5 Decomposition wavelet function ψ

The simulated results of bior6.8 wavelet transform applied on Graz dataset has been shown in Fig. 6 and Fig. 7

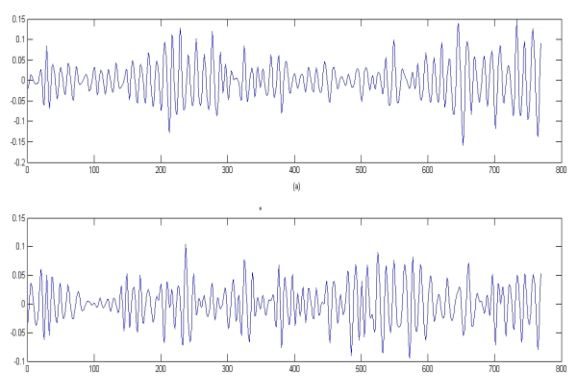

Figure 6 Wavelet coefficients Of Left signal (a) electrode C3 (b) electrode C4

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 [11]. 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 [10]. 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 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.

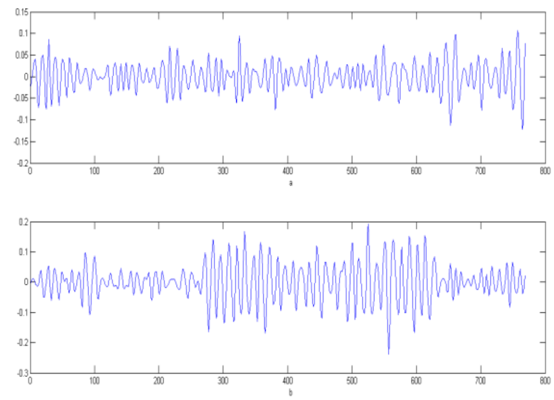

Figure 7 Wavelet coefficients Of Right signal (a) electrode C3 (b) electrode C4

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 have 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. Classification

Now the final step was to classify the signal feature sets into their particular classes with maximum accuracy. 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) with SOM based neural networks.

Linear Discriminant Analysis

LDA's main functionality is to provide the feature's spread of two sets normal with similar covariance matrix [12]. 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. Fig 8 shows the example of LDA.

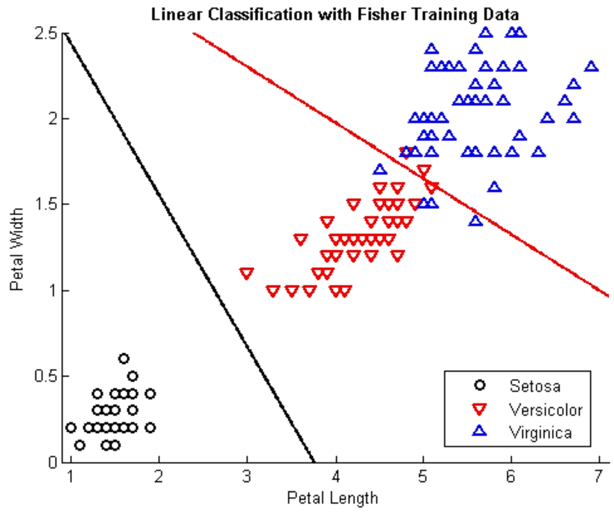

Figure 8 Example of LDA

Quadratic Discriminant Analysis (QDA)

In comparison, QDA is a comprehensive form of LDA, on condition of two classes and the groups are normally dispersed [12].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 Fig 9.

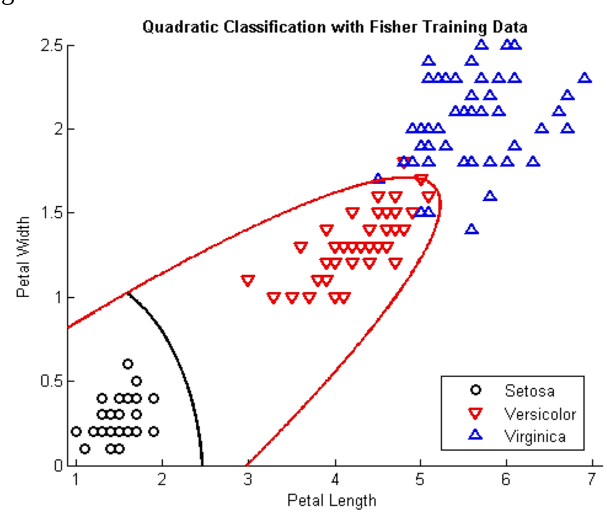

Figure 9 Example of QDA

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 apply to both untransformed and non-linear transformed feature sets [13, 14]. SVM makes a maximal dividing hyper plane with a maximum threshold amongst the groups; by increasing the dimensionality of feature space as depicted in Fig. 10.

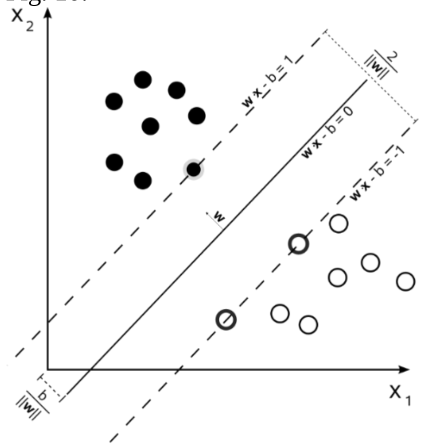

Figure 10 SVM Example

Consider a training set X defined as *{x_i, i=1, 2 ....., n}* belongs to one of the two classes ω_1 and ω_2 with corresponding labels y_i=±1. The function *xx* is known to be the discriminant function where, is the weight of coefficient vector, and 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:

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

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

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

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

Thus we have,

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

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

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. Fig 11. 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 [15].

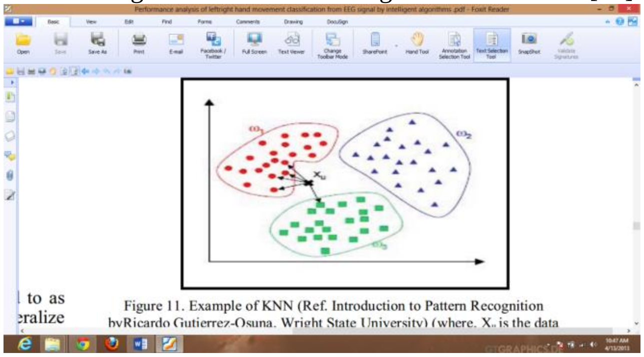

Figure 11 Example of K-Nearest Neighbour

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.

5. Performance Analysis

The both features vectors have been provided to the above mention classification algorithms using MATLAB. The classification results of both reduced and non-reduced feature vectors have been shown in Figure 12.

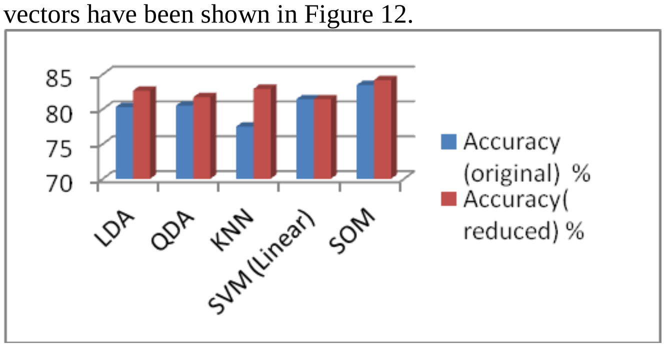

Figure 12 Result of Classification

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. CONCLUSIONS

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. The features include 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 discussed algorithms [10, 16, 17] which is also authenticated in many writings [18, 19, 20].In most of the cases, the classification of the reduced feature set of PCA has increased as compared 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%. The SOM based approach has been presented is relatively new, robust and adaptive as compared to other discussed, so in order to drive EEG sourced BCI devices (mobile robot) it has been a good approach which requires less computation and gives maximum efficiency. Our future plan is to 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.

REFERENCES

1. McFarland, Dennis J., and Jonathan R. Wolpaw. "Brain-computer interface operation of robotic and prosthetic devices." *Computer* 41.10 (2008): 52-56.
2. Daly Janis J, Wolpaw Jonathan R , “Brain–computer interfaces in neurological rehabilitation*”, Lancet Neurol* 2008; 7: 1032–43
3. Baig M.Z., Mehmood Y., Ayaz Y., “A BCI System Classification Technique using Median Filtering and Wavelet Transform”, *Log Dynamics International Conference on* , February 2014, in press.
4. 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.
5. Vidaurre C., Kramer N., Blankertz B., Schlogl A. “Time Domain Parameters as a feature for EEG-based Brain Computer Interface.” *Neural Networks 22*, 2009, 1313-1319
6. Mahmoudi B., Erfanian A. “Single channel EEG based prosthetic hand grasp control for amputee subjects.” *Proceedings of the Second Joint EMBS/BMES Conference, Houston, Texas, USA, October* 2002
7. James L. Noyes.Artificial Intelligence with Common Lisp: *Fundamentals of Symbolic and Numeric Processing. D.C. Heath, Lexingto*n, MA; 1992.
8. 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.
9. 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.
10. 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.

11. 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
12. Srivastava Santosh, Gupta Maya R., Frigyik B´ela A., "Bayesian Quadratic Discriminant Analysis" *Journal of Machine Learning Research 8* (2007) 1277-1305
13. 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.*
14. Jakkula Vikramaditya, "Tutorial on Support Vector Machine (SVM)".
15. Gunn Steve R.,"Support Vector Machines For Classification and Regression", Technical report, University of Southampton, May 1998.
16. 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*
17. 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.*
18. 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
19. Wang Y., Makeig S. "Predicting Intended Movement Direction using EEG from Human Posterior Parietal Cortex." HCI international, San Diego CA., July 2009.
20. 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
21. processing and control 4, 2009, 236-246