

# Performance Analysis of Left/Right Hand Movement Classification from EEG Signal by Intelligent Algorithms

Saugat Bhattacharyya<sup>1,a</sup>, Anwesha Khasnobish<sup>1,b</sup>, Amit Konar<sup>2,c</sup>, D.N Tibarewala<sup>1,d</sup>, Atulya K. Nagar<sup>3,e</sup>

<sup>1</sup>School of Bioscience and Engineering, <sup>2</sup>Department of Electronics and Telecommunication, <sup>3</sup>Dept. of Math & Computer Science  
<sup>1,2</sup>Jadavpur University, <sup>3</sup>Liverpool Hope University  
<sup>1,2</sup>Kolkata-700032, India, <sup>3</sup>United Kingdom

Email address: <sup>a</sup> saugatbhattacharyya@gmail.com, <sup>b</sup> anweshakhasno@gmail.com, <sup>c</sup> konaramit@yahoo.co.in,  
<sup>d</sup> ju.biomed@gmail.com, <sup>e</sup> nagara@hope.ac.uk

**Abstract**— Brain Computer interfaces (BCI) has immense potentials to improve human lifestyle including that of the disabled. BCI has possible applications in the next generation human-computer, human-robot and prosthetic/assistive devices for rehabilitation. The dataset used for this study has been obtained from the BCI competition-II 2003 databank provided by the University of Technology, Graz. After pre-processing of the signals from their electrodes (C3 & C4), the wavelet coefficients, Power Spectral Density of the alpha and the central beta band and the average power of the respective bands have been employed as features for classification. This paper presents a comparative study of different classification methods including linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA), k-nearest neighbor (KNN) algorithm, linear support vector machine (SVM), radial basis function (RBF) SVM and naive Bayesian classifiers algorithms in differentiating the raw EEG data obtained, into their associative left/right hand movements. Performance of left/right hand classification is studied using both original features and reduced features. The feature reduction here has been performed using Principal component Analysis (PCA). It is as observed that RBF kernelised SVM classifier indicates the highest performance accuracy of 82.14% with both original and reduced feature set. However, experimental results further envisage that all the other classification techniques provide better classification accuracy for reduced data set in comparison to the original data. It is also noted that the KNN classifier improves the classification accuracy by 5% when reduced features are used instead of the original.

**Keywords**- BCI, EEG, ERD/ERS, Wavelet, PSD, PCA, LDA, QDA, KNN, SVM, Bayesian

## I. INTRODUCTION

Brain computer interfaces are the new and most interesting era of research of the current century with immense potentials to improve human lifestyle. It also has the capability to bring the disabled people back to the mainstream of life providing assistive rehabilitation. With proper research and implementation of the advances BCI can contribute to the wide spectrum of areas viz. robotics, mass communication,

healthcare, military purpose, automobiles, games and entertainment, etc [1, 2].

BCI can be divided into four broad components, viz., recording of the neural activity; extraction of the intended action from that activity; generation of the desired action with prosthetic effectors; and feedback, either through intact sensation, such as vision, or generated and applied by the prosthetic device [3,4,5,6].

The brain activities for BCI can be measured using EEG (electroencephalography), ECoG (electrocorticography), fMRI (functional magnetic resonance imaging), MEG (magnetoencephalography), LPF (local field potential) [7]. EEG is the most widely used brain signal for driving a BCI system, which is a sum of the postsynaptic potentials recorded from the scalp. EEG based BCI is preferable because of its non-invasive nature, and its superior temporal resolution. EEG has an inferior spatial resolution but it can be coupled with fMRI to overcome this limitation. EEG can provide high temporal resolution in milliseconds, reflecting brain dynamic changes.

From the sensory motor cortex of the many oscillations that originate, the rolandic mu rhythm (7-13 Hz) and central beta rhythm (above 13 Hz) are relevant for BCI research [8]. In motor imagination and execution, the sensory stimulation are caused due to the neurochemical changes in the brain, the changes in the synaptic interactions or due to the changes in the intrinsic membrane properties of local neurons.

Event related synchronization (ERS) in the gamma band and event related desynchronization (ERD) in the mu and beta bands of the EEG originates in our brain during imagination or execution of body part movements. The gamma ERS and the mu-beta ERD occurs at the contralateral side of the brain near somatosensory and motor cortex area during particular limb movement. In case of ERS the power of the gamma component increases, while in case of ERD the power of the mu-beta component decreases [9, 10, 11, 12, 13 14].

Various types of electrode systems viz. 60 channel, 64 channel, 128 channel systems etc. are used for signal acquisition. But for analyzing the brain signals for motor

control related BCI, researchers have used C3, C4 and Cz as these electrodes are placed on the scalp above the motor cortex area associated with control of voluntary movements.

Extraction of prominent features describing discriminative signal properties is the main step after recording of the cortical brain signals. Feature extraction [15, 16] for required specific movement is one of the main prerequisites for the development of efficient BCI. Fast Fourier transforms (FFT), power spectral density (PSD), wavelet transform (WT), Fisher distance, short time fourier transform (STFT) [35], Hjorth parameters [36], standard deviation (SD), variance [37] are generally considered as the features. Selection of suitable features associated with the EEG signal during motor execution and, or imagination is necessary for the performance improvement of the classifiers and hence the whole system. Unwanted features would contribute to increase in computational cost and time, also resulting in improper classification results.

As the number of features extracted from the EEG signal through various feature extraction technique, forms a very large feature vector which is hard to analyze altogether. To overcome this difficulty feature reduction is done to select the most important and independent features. It is mainly done by principal component analysis (PCA) or independent component analysis (ICA). Reducing the number of features and selecting the most important features reduces the complexity of the analysis procedures, hence increasing the rate to computation with better accuracy.

Classification is the step followed by feature extraction and feature reduction to classify the movement related EEG signal into corresponding movements. Classification of the movement related synchronisation and desynchronisation from the background EEG and differentiating the movement into left right category is required for the actuation of motor controlled BCI. Performance of the BCI system mostly depends on the performance of the classifiers used. The classifier needs to be very fast for the online classification of the EEG signal. Different types of classifiers such as LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), ERNN (Elman Recurrent Neural Net), LVQ (Linear Vector Quantization), BPNN (Back Propagation Neural Net), etc are used.

The aim of this paper is to present a comparative study of different classification methods including linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA), k-nearest neighbour (KNN) algorithm, linear support vector machine (SVM), radial basis function (RBF) SVM and naive Bayesian classifiers algorithms in differentiating the raw EEG data obtained, into their associative left/right hand movement. Performance of left/right hand classification is studied using both original features and reduced features. The feature reduction here has been performed using Principal component Analysis (PCA). Using independent t-test the most significant feature is selected from the total feature set. The features considered in this paper include wavelet coefficients, average band power and power spectral density. The raw EEG signal is preprocessed by noise removal using appropriate filters prior to the analysis.

The paper is divided into six sections. The organization of the experimental data and data preprocessing is explained in section II. The feature extraction principle is briefly outlined in section III. Motor imagery classifiers (LDA, QDA, KNN, linear SVM, RBF SVM and naive Bayesian classifier) are described in section IV. Performance analysis of the classifiers is given in section V. Experimental results and conclusions are listed in section VI.

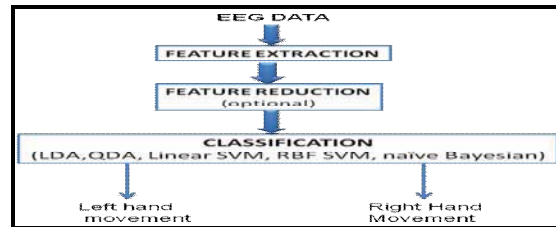


Figure 1. Proposed BCI scheme.

## II. EXPERIMENTAL DATA DESCRIPTION

The experimental data was obtained from BCI Competition 2003 provided by Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz. This dataset was recorded from a normal subject (female, 25 yr) during a feedback session where the subject was made to sit in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery left-right movement in which the order of left and right cues were random. The recording was made using a G.tec amplifier and an Ag/AgCl electrode and three bipolar EEG channels were measured over C3, Cz and C4 electrode (Fig.2). The experiment consists of 7 sessions with 40 trials each conducted on the same day with several minutes break in between. In each trial, the first 2 seconds was quite. In the 2<sup>nd</sup> second an acoustic stimulus indicates the beginning of the trail with a fixation cross '+' displayed on the screen and at the 3<sup>rd</sup> second the visual cue (left-right arrow) is displayed. At the same time the subject was asked to move the bar in the direction of the cue as feedback. The feedback was based on AAR parameters of channel C3 and C4 and the AAR parameters were combined with a discriminant analysis into one output parameter (Fig.3). The EEG data was sampled at 128Hz.

A total of 280 trials were given of 9 second each. Out of the three electrodes used, C3 and C4 are selected for this study. Cz is left out because it is of little relevance for extracting information on left-right movement [17]. The trials for training and testing were selected randomly to prevent any systematic effect due to feedback. So, a total of 140 trials were selected for training and the rest 140 trials for test. As the visual cue started from t=3 sec to t=9 sec, thus, only the data for this time interval was selected. Now it is known that the brain electrical activities mainly occur in the 0.3-40Hz bands, and the higher frequencies can be considered as noise based on their environments and recording techniques. Thus a band-pass filter is used to filter in the frequency band: 0.5-30 Hz.

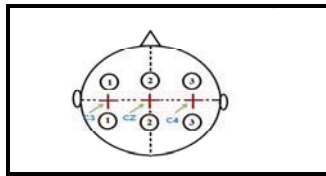


Figure 2. Electrode placement based on the experiment

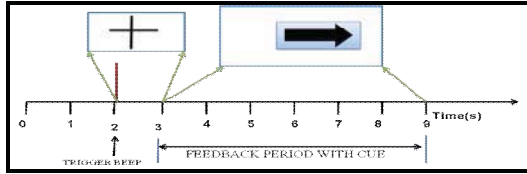


Figure 3. Timing scheme of the experiment

### III. DESCRIPTION OF THE FEATURES EXTRACTED

#### A. Wavelet Features Feature Extraction

Wavelet transforms is a very effective way to extract features from an EEG signal [18, 19]. Their ability to discriminate both the temporal and spectral domain features of the signals makes them an important asset for EEG analysis. Also the wavelet transform do not suffer from the time-frequency trade off inherent in Short Time Fourier Transform (STFT) and Fourier Transform (FT) as their multi-scale approximation allows for effective localization of the signal with various spectral-temporal characteristics. Thus for a non-stationary signal like EEG, it is an effective analysis tool. The discrete wavelet transforms analyzes the signals at different resolutions by decomposing the signal into coarse approximation and detail information. Each level comprises of two digital filters and two down-samplers by 2. The down-sampled outputs of the first high-pass and low-pass filters provides the detail D1 and approximation A1, respectively. The first approximation is further decomposed and the process continued, until the desired result is obtained [20, 21].

In the present study, Daubechies (db) mother wavelet of order 4 is used. After trials with the EEG data, the D3 features (7.81-15.62 Hz) i.e., the difference of the third level coefficient for the respective electrodes were selected as one of the feature components for the final feature vector (C4-C3). Figure 4 and 5 shows the wavelet decomposition for left-right imagery for C3 and C4 electrode.

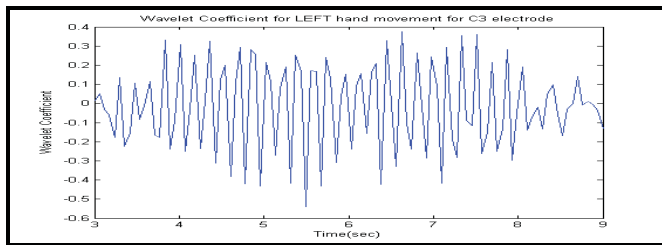


Figure 4a. Wavelet Coefficient for Left movement for C3 electrode

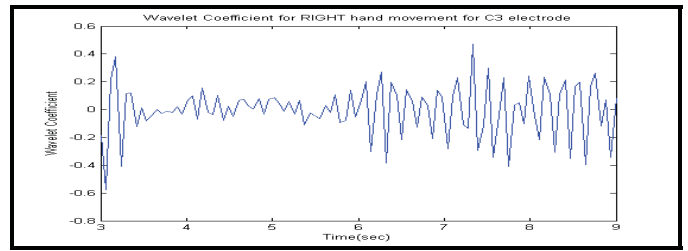


Figure 4b. Wavelet Coefficient for Right movement for C3 electrode

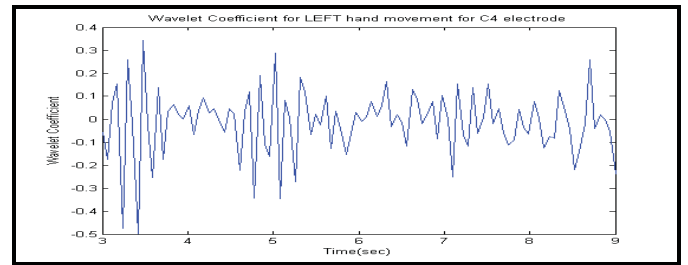


Figure 5a. Wavelet Coefficient for Left movement for C4 electrode

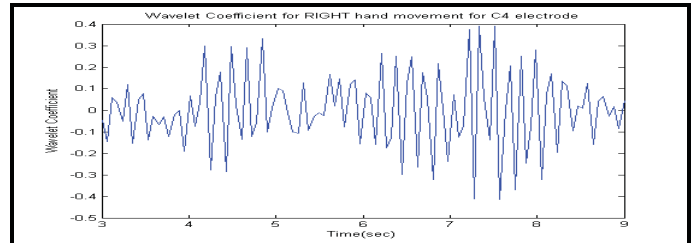


Figure 5b. Wavelet Coefficient for Right movement for C4 electrode

#### B. Spectral Estimation Method

Spectral density methods extract information from a signal to describe the distribution of its power in the frequency domain. The power spectral density (PSD) is defined as the Fourier transform (FT) of the signal's autocorrelation function provided that the signal is stationary in a wide sense [21]. Thus for an EEG signal segmenting the complete time series data would be an ideal approach.

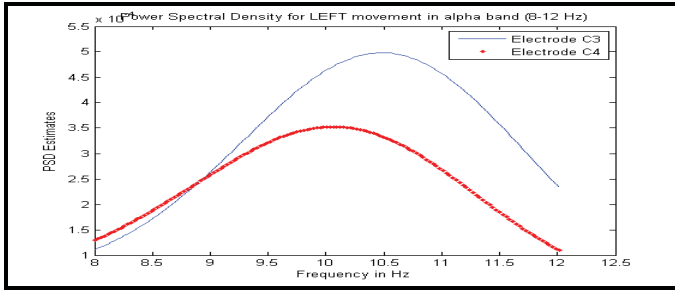
For this paper, the Welch approach was applied along with a Hamming window of length 64. The Welch method divides the times series data into overlapping segments, computing a modified periodogram of each segment and then the PSD estimates is averaged. The PSD estimates were obtained for the frequency band of 8-25 Hz which comprises of both the alpha or mu band (8-12Hz) and the central beta band (18-25Hz) for each respective electrode. Also the average power was obtained for each band. Then the difference of the PSD estimates (formula 1) and average power (formula 2) is selected as another feature for this study.

$$F_{PSD} = \sum_{f=a}^b PSD_{C4}(f) - \sum_{f=a}^b PSD_{C3}(f) \quad (1)$$

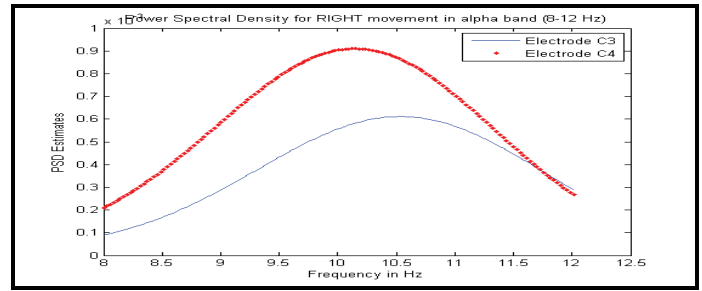
$$F_{POW} = POW_{C4} - POW_{C3} \dots \dots \dots (2)$$

where,  $PSD_{C3/C4}$  is the PSD estimates of the respective electrodes in [a, b], where a & b is the frequency range,  $POW_{C4/C3}$  is the average power of the respective electrodes in [a, b]. Figure 6 and 7 shows the PSD estimates for left-right

imagery for C3 and C4 electrodes, for both alpha and beta band. Figure 8 shows the PSD estimates for left-right imagery for the difference of the two electrodes.

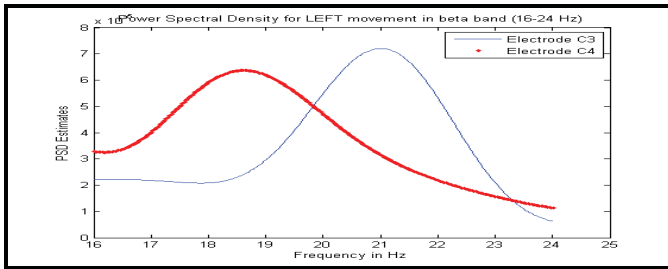


6(a)

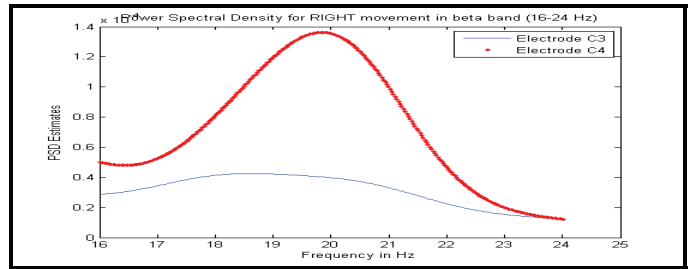


6(b)

Figure 6. Alpha Band PSD estimate for a) left movement b) right movement

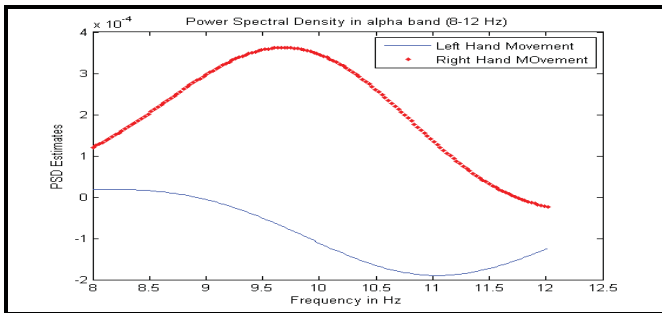


7(a)

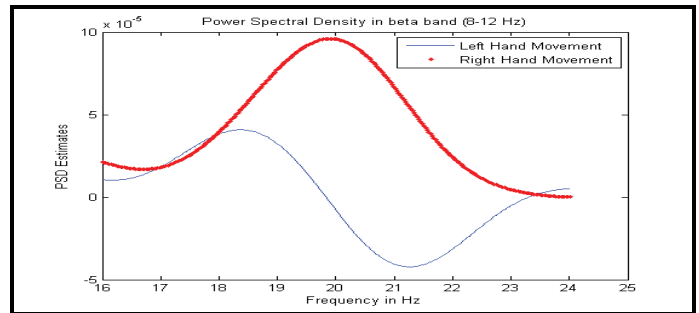


7(b)

Figure 7. Beta Band PSD estimate for a) left movement b) right movement



8(a)



8(b)

Figure 8. Power Spectral Density of a) alpha band b) beta band for the difference of the two electrodes C3 and C4 for left/right hand movement

#### IV. CLASSIFIERS

##### A. Discriminant Analysis

###### 1) Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis (LDA) is a Bayes optimal classifier, provided the distribution of features in each of two classes is normal with the same covariance matrix [22]. LDA projects d-dimensional data into a line, reducing the dimensionality by mapping L distributions to (L-1) dimensional subspace. LDA maximizes the ratio of between-class variance to the within-class variance in any particular data set, thus providing maximal separability. LDA doesn't

change the location of the original data sets but provides more class separability and draws a decision region between the given classes. The LDA finds a one-dimensional subspace in which the classes are usually well separated by a linear separating hyper plane. The discriminant function [23] is given by

$$d_k(x) = 2\mu_k^T \sum_k^{-1} X - \mu_k^T \sum_k^{-1} \mu_k - 2\log \pi(k). \quad (3)$$

where,  $k$ =class,  $X$  is the set of measurements,  $\mu_k$  is the mean vector,  $\pi_k$  is the prior probability and  $\Sigma_k$  is the covariance matrix.

2) *Quadratic Discriminant Analysis(QDA)*

QDA is a generalized version of LDA, provided there are only two classes of points and the measurements are normally distributed [24]. However unlike LDA, the assumption that the covariance of each class is identical is not taken into consideration in QDA. Further, the surface that separates the subspaces will be a conic section (like parabola, hyperbola, etc.). When the equation (3) is multiplied by -2, the discriminant function is given by

$$d_k(X) = (X - \mu_k)^T \Sigma_k^{-1} (X - \mu_k) + \log |\Sigma_k| - 2 \log \pi_k \quad (4)$$

and the discriminant rule is given by

$$d_{\hat{k}}(X) = \min_{1 \leq k \leq K} d_k(X) \Leftrightarrow \max_{1 \leq k \leq K} p(k/x) \quad (5)$$

where,  $p(k/x)$  =posterior distribution. Using this rule is called the QDA. QDA generally requires larger sample size than LDA.

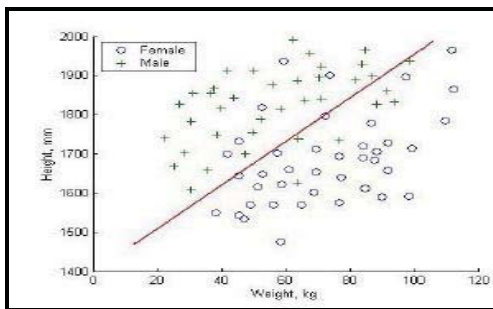


Figure 9. Example of linear discriminant analysis

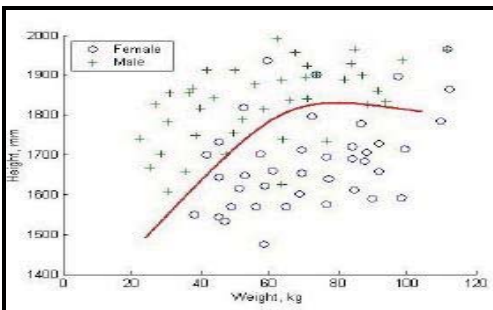


Figure 10. Example of quadratic discriminant analysis

B. *K- Nearest Neighbour*

The  $k$ - nearest neighbor algorithm is also referred to as Lazy Learning algorithm as it defers the decision to generalize the training data points till a new query is encountered. The KNN is based on the assumption that class probabilities are locally approximately constant, so it is amongst the simplest of

all machine learning algorithms. The  $k$ -nearest neighbors (KNN) algorithm finds the  $k$ -nearest neighbors among the training set, and the categories of the  $k$ -nearest neighbors are used to weigh the category candidates.  $k$  is a positive integer, which is typically small. The distance is calculated using different methods, such as Euclidean distance, Mahalanobis distance, Minkowski distance, Manhattan distance etc. The objects are represented in multi dimensional feature space by position vectors in order to identify the neighbors. The performance of this algorithm greatly depends on two factors: a suitable similarity function and an appropriate value for  $k$ . If  $k$  is too large, big classes will overwhelm the small classes whereas if  $k$  is too small, the advantage of KNN algorithm cannot be traced. Equations (6) and (7) are the widely used strategies for this algorithm [25].

$$y(d_i) = \arg \max_k \sum_{x_j \in kNN} y(x_j, c_k) \quad (6)$$

$$y(d_i) = \arg \max_k \sum_{x_j \in kNN} Sim(d_i, x_j) y(x_j, c_k) \quad (7)$$

where,  $d_i$  is a test document,  $y(x_j, c_k) \in \{0, 1\}$  indicates whether  $x_j$  belongs to class  $c_k$ ,  $Sim(d_i, x_j)$  is the similarity function for  $d_i$  and  $x_j$ . Equation (6) shows that the prediction will be the class that has the largest number of members in the  $k$  nearest neighbor. Equation (7) indicates the class with maximal sum of similarity will be the best result.

KNN algorithm is affected by local structures of data, sensitive to noise, requires large memory and is computationally expensive. When  $k$  is smaller, variance is higher making the system less stable. When  $k$  is larger, bias is higher which reduces the precision. In the training phase of the KNN algorithm feature vectors and class labels of the training data points are stored. The test data points are represented as vectors in the feature space during the classification phase. The distance is calculated from all stored vectors to the new vectors and  $k$  closest samples are selected. The new vector is classified to a particular class most commonly by predicting the new vector to the most common class among the  $k$  nearest neighbours. KNN is nearly optimal for large sample limits ( $n \rightarrow \infty$ ) but it is highly susceptible to curse of dimensionality.

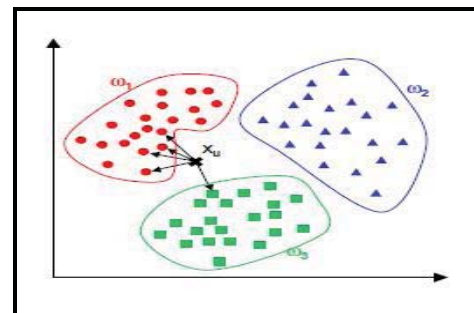


Figure 11. Example of KNN (Ref. Introduction to Pattern Recognition by Ricardo Gutierrez-Osuna, Wright State University) (where,  $X_u$  is the data point with unknown class label,  $k=5$  neighbours of which 4 belong to  $\omega_1$  and belong to  $\omega_3$ . As  $\omega_1$  is the predominant class  $X_u$  is assigned to  $\omega_1$ )

C. Support vector machine (SVM)

A support vector machine classifies by construction of a N-dimensional hyperplane that optimally separates the data in two categories. SVM is flexible in choosing similarity function [26]. The separating hyper plane is specified only by the support vectors. As the complexity does not depend on the dimensionality of the feature space, SVM can deal with large feature spaces. SVM is sensitive to noise and it can consider only two classes. A row of predictor values or a set of features that describes one case is called the vectors. Support vectors are the most important training points that define the hyper plane. The aim to classification by SVM algorithm (Fig. 12) is to find an optimal hyperplane that separates clusters of vector in such a way that the vectors lie on the either side of the hyperplane belonging to two different non-overlapping classes.

In classification by SVM the training is relatively easy. It scales relatively well to high dimensional data and the trade-off between classifier complexity and error can be controlled explicitly. The limitations of SVM are the choice of kernels, speed and size (in both training and testing). SVMs are characterized by the choice of its kernels.

1) Linear SVM

The linear classifier relies on dot product between vectors  $K(x_i, x_j) = x_i^T x_j$ , where  $x_i$  and  $x_j$  are the support vectors[27]. The simplest way to divide the vectors in two classes is with a straight line, flat plane or N-dimensional hyperplane. The training data  $\{x_i, y_i\}$ ,  $i = 1, 2, \dots, L$ ,  $y_i \in \{-1, 1\}$ ,  $x_i \in R^d$ . The points  $x$  which lie on the hyperplane satisfy  $w \cdot x + b = 0$ , where

$w$  is normal to the hyperplane,  $\frac{|b|}{\|w\|}$  is the perpendicular distance from the hyperplane to the origin, and  $\|w\|$  is the Euclidean norm of  $w$ . Let  $d^+$  ( $d^-$ ) is the shortest distance from the separating hyperplane to the closest positive (negative) example. The “margin” of a separating hyperplane is given by  $(d^+ + d^-)$ . For the linearly separable case (Fig. 12), the support vector algorithm simply looks for the separating hyperplane with largest margin. This can be formulated as follows: suppose that all the training data satisfy the following constraints:

$$x_i \cdot w + b \geq +1 \text{ for } y_i = +1 \dots \dots \dots (8)$$

$$x_i \cdot w + b \leq -1 \text{ for } y_i = -1 \dots \dots \dots (9)$$

These can be combined into one set of inequalities:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \dots \dots \dots (10)$$

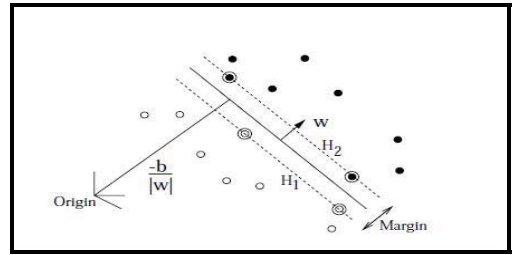


Figure 12. Linear separating hyperplane. The support vectors are encircled.

2) Radial basis function (RBF) SVM

When the vectors are separated by non-linear region, the SVM uses a kernel function to map the data into a different space where a hyperplane can be used for separating the vectors. Certain function that corresponds to an inner product in some expanded feature space is referred to as kernel function. According to Mercer’s theorem, every semi positive definite symmetric function is a kernel. Kernel function transforms the data into higher dimensional space to make it possible for the separation of the vectors (Fig. 13). The dot product becomes  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  when every data point is mapped into high-dimensional space via some transformation  $\Phi: x \rightarrow \phi(x)$ . The kernel matrix,  $K_{ij} \equiv K(x_i, x_j)$ , is a Gram matrix (a matrix of dot products (Horn, 1985)) in  $H$  (i.e. the Euclidean Space) [28]. It is necessary to choose  $l$  training points such that the rank of the matrix  $K_{ij}$  increases without limit as  $l$  increases. The radial basis function is given by

$$K(x', x) = \exp \left[ - \frac{\|x' - x\|^2}{\sigma^2} \right],$$

which defines a spherical kernel where  $x'$  is the centre and  $\sigma$ , supplied by the user defines the radius.

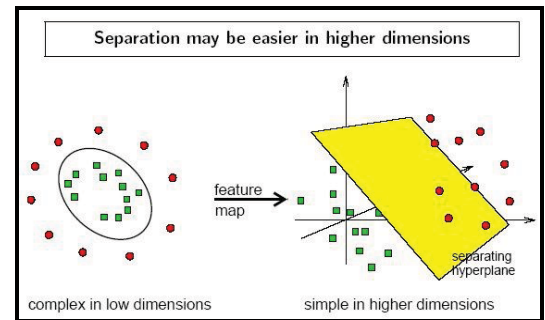


Figure 13. Transformation of data into higher dimensional space by kernel function.

D. Naive Bayesian Classifier

Bayesian classifiers are the statistical classifiers predicting class membership probabilities [29]. A Bayesian approach splits this posterior distribution into a prior distribution  $P(C)$  and likelihood  $P(D|C)$ :

$$\arg \max_c P(C = c | D = d) = \frac{\arg \max_c P(D = d | C = c) P(C = c)}{P(D = d)} \dots \dots \dots (11)$$

The denominator  $P(D = d)$  is a normalizing factor that can be ignored when determining the maximum a posteriori class, as it does not depend on the class. The term  $P(D = d|C = c)$  is the likelihood of the given description of the class (abbreviated as  $P(d|c)$ ). Estimates of likelihoods of the training data are computed by the Bayesian classifiers. The assumption that all attributes are independent given the class solves the problem of joint probability distribution:

$$P(A_1 = a_1, \dots, A_n = a_n | C = c) = \prod_{i=1}^n P(A_i = a_i | C = c) \dots \dots (12)$$

This assumption is called naive Bayesian assumption which is utilized by naive Bayesian classifiers. Thus naive Bayesian classifier ignores interaction between attributes within individuals of the same class. Number of terms to be estimated from training data is much smaller : ( number of distinct attribute values)  $\times$  (number of possible target values). The naive Bayesian classifier simplifies the computation involved.

### V. PERFORMANCE ANALYSIS

From the total raw data, only the data from the time interval of t=3 to 9s are taken. From the total of 280 trials, 140 were selected as training data and the rest 140 as test. All of the data were bandpass filtered between the frequency ranges of 0.5-30Hz. From the two electrodes of interest, namely, C3 and C4, wavelet coefficients, PSD estimates for the alpha and beta bands and their corresponding powers were selected as the features for this study using the Wavelet toolbox and Signal Processing Toolbox in MATLAB (Table I).

TABLE I. FEATURE VECTORS WITH THEIR RESPECTIVE SIZE

FEATURE VECTORS	SIZE
	<i>(No. of Features per Trial <math>\times</math> No. of Trials)</i>
Wavelet Coefficient (D3)	102 $\times$ 140
Power Spectral Density (for 8-25 Hz)	768 $\times$ 140
Average Power	1 $\times$ 140
<b>Total</b>	<b>871 <math>\times</math> 140</b>

On the total feature set, independent sample t-test was performed to select the most significant features required for proper classification, in which a significant level of 5% was used. These data was labeled as non-reduced features (500  $\times$  140).

Each of the features was then reduced using Principal Component Analysis (PCA) to around one-tenth of their original size (except the average power). So, the total size of the reduced feature vectors becomes 91 $\times$ 140. The original feature vector and the reduced feature vector were fed into LDA, QDA, KNN, SVM and naïve Bayesian in the MATLAB environment (CLASSIFY, KNNCLASSIFY,

SVMCLASSIFY, etc). The accuracy obtained for both the sets of feature vectors for their respective classifiers are shown in Table 2.

TABLE II. RESULT OF CLASSIFICATION

Serial No.	Classifiers	Accuracy (of non-reduced features) in %	Accuracy (of reduced features: PCA) in %
1	LDA	79.29	81.14
2	QDA	79.29	80.71
3	KNN	76.43	81.43
4	SVM (Linear)	80.71	80.71
5	SVM (RBF kernel)	82.14	82.14
6	Naïve Bayesian	80	81.43

It is as observed from Table II for both the sets of feature vectors, RBF kernelized SVM classifier indicates the highest performance accuracy of 82.14%. However, for all other classifiers there is an increase in performance accuracy for the reduced feature set as compared to the original set, which indicates the increased contribution of the principal components in the classification. Also, KNN classifier shows the most significant increase in the performance accuracy from 76.43% to 81.43%.

### VI. CONCLUSION

In this paper, features are extracted from the preprocessed EEG signal and fed to the motor imagery classifiers for differentiating the EEG signal to its corresponding left-right limb movement. Wavelet transform, power spectral density estimate and average band power estimates are techniques followed in this study for feature extraction. An extensive study of comparative performance analysis of the different classifiers (including LDA, QDA, KNN, linear SVM, RBF SVM and naïve Bayesian) are presented in this paper. Two approaches were used in this paper: in one sample t-test was performed on the original feature set while in the other the original feature set was reduced using PCA. It is evident from the result that kernelized SVM gave the best performance as compared to other methods in our previous works [30, 31] which is also validated in many literatures [32, 33, 34]. Also in many cases, the performances of PCA based features were better than the non reduced features due to the increased contribution of the principal components, which simultaneously increases the classification accuracy. The processing of the features requires further validation and study to improve the accuracy of the classifiers. Also, it is held that the combination of feature vector is vital for proper classification, thus newer features need to be tried out to further improve the classification of left-right motor imagery. Our approach of feature extraction and classification presented here is very simple and robust. To control EEG based BCI devices it is required to find out more relevant features with less computational time and with higher computational

efficiency. Future study in this direction will aim at techniques for optimizing feature selection, extraction and classification methodologies to be implemented in online classification of EEG data for BCI research.

#### REFERENCES

- [1] McFarland Dennis J., Wolpaw Jonathan R. "Brain-Computer Interface Operation of Robotic and Prosthetic Devices", IEEE Computer Society, 2008
- [2] Daly Janis J, Wolpaw Jonathan R , "Brain-computer interfaces in neurological rehabilitation", *Lancet Neurol* 2008; 7: 1032-43
- [3] Schwartz A.B., Cui X.T., Weber D.J., Moran D.W. "Brain Controlled Interfaces: Movement Restoration using Neural Prosthetics." *Neuron* 52, October 2006, 205-220
- [4] Lebedev M.A., Nicolelis, "Brain-machine interface: Past, present and future", *Trends Neurosci.* 29(9), September 2006, 536-546
- [5] Vaughan T.M., Heetderks W.J., Trejo L.J., Rymer W.Z., Weinrich M., Moore M.M., Kubler A., Dobkin B.H., Birbaumer N., Donchin E., Wolpaw E.W., Wolpaw J.R. " Brain computer interface technology: A review of the second international meeting", *IEEE Trans. Neural Syst. Rehab. Eng.* 11(2), June 2003, 94-109
- [6] Wolpaw J.R., Birbaumer N., Heetderks W.J., McFarland D.J., Peckham P.H., Schalk G., Donchin E., Quatrano L.A., Robinson C.J., Vaughan T.M. " Brain computer interface : A review of the first international meeting", *IEEE Trans. Rehabilitation Eng.* 8(2), June 2000, 164-173
- [7] Anderson R.A., Musallam S., Pesaran B., "Selecting the signals for a brain-machine interface", *Curr Opin Neurobiol* 14(6), December 2004, 720-726
- [8] Pfurtscheller G., Neuper C., "Motor imagery activates primary sensorimotor area in humans", *Neuroscience Letters* 239, December 1997, 65-68
- [9] Pilla Jr, V. Lopes H.S., "Detection of movement related Desynchronization of the EEG using neural networks", *Proceedings of the 22<sup>nd</sup> Annual EMBS International Conference, Chicago, July 2000*
- [10] Qiang C., Hu P., Huanqing F., " Experiment study of the relation between motion complexity and event related Desynchronization/synchronization", *2005 First International Conference on Neural interfaces and Control Proceedings, China, May 2005*
- [11] Matsunaga T., Katayama Y., Hayami T., Iramina K. "Measurement of mu/beta ERD and gamma ERS during the imagination of body parts movement." *30<sup>th</sup> Annual International IEEE EMBS Conference Vancouver Canada, August 2008*
- [12] Hema C.R., Paulraj M.P., Yaacob S., Adom A.H., Nagarajan R. "Recognition of motor imagery of hand movements for a BMI using PCA features." *2008 International Conference on Electronic Design, Penang, Malaysia, December 2008*
- [13] Thomas K.P., Guan C., Tong L.C., Prasad V.A. "An adaptive filter bank for motor imagery based brain computer interface." *30<sup>th</sup> Annual International IEEE EMBS Conference, Vancouver, Canada, August 2008*
- [14] 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
- [15] Cososchi S., Strungaru R., Ungureanu A., Ungureanu M., " EEG feature extraction for motor imagery", *Proc. of IEEE Engineering in Medicine and Biology Society, August 2006, 1142-1146*
- [16] Yom-Tov E., Inbar G.F., "Feature selection for the classification of movements from single movement-related potentials", *IEEE Trans. Neural Syst. Rehab. Eng.* 10, 2002, 170-178
- [17] Xu Huaiyu, Lou Jian, Su Ruidan, Zhang Erpang "Feature Extraction and Classification of EEG for imaging left-right hands movement."
- [18] Schloegl A., "Dynamic spectral analysis based on an autoregressive model with time-varying coefficients", *IEEE-EMBC and CMBEC, 1995*
- [19] Darvishi S., Al-Ani A. "Brain-computer interface analysis using continuous wavelet transform and adaptive neuro-fuzzy classifier", *Proc. 29<sup>th</sup> Int. Annu. Conf. IEEE Eng. Med. Biol. Soc., August 2007, pp. 3220-3223*
- [20] Xu Q., Zhou H., Wang Y., Huang J. "Fuzzy support vector machine for classification of EEG signals using wavelet based features." *Medical Engineering & Physics* 31, 2009, pp. 858-865
- [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] Balakrishnama S., Ganapathiraju A. "LINEAR DISCRIMINANT ANALYSIS - A BRIEF TUTORIAL"
- [23] Daniella Birkel, "Regularized Discriminant Analysis"
- [24] Srivastava Santosh, Gupta Maya R., Frigiyk 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 20<sup>th</sup> 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] BURGES CHRISTOPHER J.C., "A Tutorial on Support Vector Machines for Pattern Recognition", *Kluwer Academic Publishers, Boston, pp- 1-43*
- [29] FLACH PETER A. , LACHICHE NICOLAS, "Naive Bayesian Classification of Structured Data", *Machine Learning, , Kluwer Academic Publishers, Boston, pp- 1-37*
- [30] Bhattacharyya S., Khasnobish A., Konar A., Tibarewala D.N., "Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data", *Accepted for oral presentation in International conference on systems in medicine and biology, IIT Kharagpur, 2010*
- [31] Khasnobish A., Bhattacharyya S., Konar A., Tibarewala D.N., "K-Nearest neighbor classification of left-right limb movement using EEG data.", *Accepted for oral presentation in International conference on Biomedical Engineering and assistive technologies, NIT Jalandhar, 2010.*
- [32] Thomas K.P., Guan C., Tong L.C., Prasad V.A. "An adaptive filter bank for motor imagery based brain computer interface." *30<sup>th</sup> Annual International IEEE EMBS Conference, Vancouver, Canada, August, 2008*
- [33] Wang Y., Makeig S. "Predicting Intended Movement Direction using EEG from Human Posterior Parietal Cortex." *HCI international, San Diego CA., July 2009.*
- [34] 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
- [35] Zhendong Mu Dan Xiao, Jianfeng Hu. "Classification of motor imagery EEG signals based on STFTs."
- [36] 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
- [37] 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*