Classification of Brain States using Subject-Specific Trained Classifiers

N. Irtiza[1], H. Farooq[2]

[1]Electrical Engineering Department, Bahria University, Karachi, Pakistan,
[2]Computer Science Department, Bahria University, Karachi, Pakistan,
[1]navinza@hotmail.com

***Abstract*-**One of the key components to develop a usable Electroencephalography (EEG) based Brain Computer interface (BCI) is the efficient classification of EEG patterns using Machine Learning classifiers. This paper presents a comparison of Linear Discriminant Analysis (LDA), Naïve Bayes and Decision Tree classifiers by applying to the EEG data. The classifiers are applied on BCI Competition III: Dataset V that consists of three cognitive tasks, namely, right and left hand imagery movement and the imagination of any word starting from a given random letter. The BCI experiments for this data have been performed with three subjects. For subjects 1 and 2, the Naïve Bayes classifier provides best results while for subject 3 the maximum accuracy is achieved from LDA classifier. In order to improve the accuracy further, it has been proposed to apply combination of classifiers based on Multiple/Ensemble Classifier System concept on data for single subject with different sessions of data recording. By combining the classifiers LDA and Decision Tree, maximum accuracies of 81%, 70% and 56% for subjects 1, 2 & 3 respectively have been achieved that are comparable with the accuracies achieved by the winner of the competition. It is concluded that instead of employing single classifier, the approach for using combination of classifiers significantly improves the performance of a BCI system.

***Keywords*-**Brain Computer Interfacing, Machine Learning, Electroencephalography, Classification, Multiple Classifier System

I. INTRODUCTION

A communication system that transmits signals from brain without using any muscular activity to a machine or a computer is termed as Brain Computer Interface (BCI) [i]. The Electroencephalography (EEG) based signal generated by a human brain performing a particular cognitive task or brain activity like imagination of hand or movement or observing several auditory or visual stimuli can be helpful to determine the intention and emotion of the user. The most common approach for designing a BCI system is to perform analysis and interpretation of EEG signals in such a manner that they vary the state of a machine [i]. The EEG signal captured from brain consists of six different oscillations having specific frequency bands. The oscillations based on frequency ranges as well as location of their origin are categorized. They are delta (less than 4 Hz), theta (4-7 Hz), alpha (8-15 Hz), beta (12- 30 Hz), mu (8-12Hz) and gamma (32+Hz). Based on BCI System, various computer based applications have been designed [ii, iii]. Motor imagery is defined as a cognitive task by which a subject simulates any specific action in her mind without actually performing the movement. Motor imagery has been widely considered as a major scenario in BCI related studies [iv-vi]. The Berlin Brain Computer Interface group in Germany has developed a BCI application named as the P300 or Hex-O speller which is a EEG signal based spelling application in which the user has to control the size and rotation of the arrow displayed on the screen using motor imagery. He has to select a cell in a hexagon that consists of 6 cells such that each cell consists of either a letter or group of letters [ii].

Efficient classification of different electroencephalogram (EEG) patterns plays instrumental role in designing a usable BCI System. The purpose of the classification stage in a BCI device is to assign automatically a class or label to the feature vector that has been previously extracted. A classifier is defined as a function that partitions a set of data or objects into two or multiple classes. Thus, in machine learning the purpose of the classification is to develop a rule based on a set of samples or observations that allocate an observation or a sample x to one of multiple classes. Here, x denotes a feature vector of N-dimensional data. For the simplest case, there can be only two different binary classes. In such case, a classifier can be formulated as a decision function $f: \mathbb{R}^N \rightarrow \{-1, +1\}$, that assigns an instance x to one of the classes, assigned by +1 and -1 [vii]. The predicted class label represents the state of user's mind while performing the BCI experiment.

In this paper, we aim to validate one of the stages of our proposed methodology in our previous work to find out the neural correlates of Neuroception [viii]. The proposed work specifically focuses on the problem of assessment of neural correlates for fear-induced EEG activity acquired while user's brain is functioning unconsciously to differentiate among the states of safety vs. dangerous. We proposed that LDA will be used to perform classification of the cognitive state of the subject if he is in state of fear or not. To validate this statement, initially the LDA classifier is applied on the dataset BCI Competition III: dataset V from Berlin BCI group. In addition to LDA, Naïve Bayes and Decision tree are applied on the same dataset for comparison. In order to improve the performance of the classifier, concept of Multiple / Ensemble Classifier has been applied and it is found that the approach has significantly improved the results. Employing the combination of LDA and Decision tree, the prediction accuracy has remarkably increased.

The rest of the paper is arranged as follows. Section II explains some BCI studies emphasizing on what feature extraction and classification techniques have been used. Section 3 consists of methodology with description of dataset and EEG feature based on Power Spectral Density (PSD). Details are given regarding how LDA, Naïve Bayes (NB) and Decision Tree (DT) classifiers are used in this work. The approach combining LDA with Decision Tree is also presented. The Section IV mentions the results obtained from each of the classifiers. Section V discusses the results and the performance with the existing results is compared. Finally, the paper is concluded in Section VI.

TABLE I
STUDIES SHOWING MACHINE LEARNING TECHNIQUES FOR BCI APPLICATIONS

Authors' work	Features Extracted	Machine learning classifiers	Performance Measure / Value
[ix]	• Power Spectrum • Wavelet • Entropy • Fractal Dimension	• LDA • PCA • SVM	Classification Accuracy / 91.77%-
[x]	• Wavelet	• Independent Component Analysis (ICA)	
[iii]	• Event Related Potential	• Linear FDA	• Offline binary classification / 77.7% • Online multiclass accuracy / 89.37%
[xii]	• Power Density	• Short term fourier transform (STFT) • Genetic Algorithm (GA)	Classification Error Rate / 20%

The rest of the paper is arranged as follows. Section II explains some BCI studies emphasizing on what feature extraction and classification techniques have been used. Section III consists of methodology with description of dataset and EEG feature based on Power Spectral Density (PSD). Details are given regarding how LDA, Naïve Bayes (NB) and Decision Tree (DT) classifiers are used in this work. The approach combining LDA with Decision Tree is also presented. The Section IV mentions the results obtained from each of the classifiers. Section V discusses the results and compared the performance with existing work. Finally, the paper is concluded in Section VI

II. RELATED WORK

In any BCI application, the features extracted from acquired EEG signals are translated into interpretable commands. For this purpose, different algorithms from machine learning have been employed for the past several years [ix. xiii]. Some of the techniques and the features extracted in BCI systems are mentioned in Table I. A study [ix] is conducted to identify the specific characteristics associated with emotions using EEG signals. Furthermore, pattern of emotion changes while the subject is experiencing during the BCI experiment are also investigated. For this purpose, different movie clips are designed to induce emotions of subjects. From EEG signals captured, three different features based on power spectrum, non-linear dynamical analysis and wavelets are extracted. Support Vector Machine (SVM) is used for classification purpose. In order to achieve reduction in features dimensionality, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied and compared. The best classification accuracy of 91.77% is achieved in this work [ix].

A computer game to find out cortical oscillatory dynamics associated with three different nature of social behavior including aggressive, avoidant and friendly is designed in [x]. In this work, EEG responses are analyszed using wavelet transform. For feature extraction and classification, the technique of Independent Component Analysis (ICA) is used.

A novel method using auditory evoked potentials for designing a multiclass spelling BCI application is proposed in [iii]. In this study, the subjects have focused their attention to different auditory stimuli varying both in pitch and direction. The signals operating the text entry application are categorized on the basis of nine different control signals. The proposed method was named as PASS2D that achieved information transfer rate (ITR) of 0.8 characters per minute. For classification purpose, fisher discriminant analysis (FDA) is used. In case of offline binary classification, 77.7% accuracy is achieved. While for online multiclass scenario, 89.37% decisions were correctly classified [iii].

Extraction of optimal feature set based on power density of EEG response is performed in [xii]. Pre-processing is conducted using spatial and temporal filters for short-time fourier transform (STFT) to extract time-frequency features. Genetic algorithm (GA) is then applied for selection of optimal features. In this work, the datasets from BCI competition III and IV are used. In addition, BCI experiments are also performed to apply the proposed approach. SVM is applied for classification and achieved the classification error rate to 20% [xii]

A. *Multiple / Ensemble Classifier System*

Multiple classifier systems (MCSs), classifier committees or ensemble classifiers perform the task of classification not based on a single classifier but based on a combined decision of a set of at least two classifiers [xvi]. Various studies have used this approach. In [xix] Support Vector Machine (SVM) and decision tree have been integrated for understanding and prediction of trans-membrane segments.

In [xx] the performance evaluation of three ensemble algorithms for EEG signal classification of tasks for mental imagery has been performed. Using k-nearest neighbor as base classifier, combined with decision tree and SVM, classification is performed on real EEG recordings. Experimental results suggest the practicability of ensemble classification methods for EEG Signals classification [xx]. In [xxi] combination of LDA, ADABoost and Decision tree classifiers to detect different cases of occluded as well as non-occluded faces is used. The ensemble of LDA and decision tree combines the outputs from each of the individual classifiers.

The two architectures used in the design of MCS are illustrated in Fig. 1. Firstly, in parallel topology, each classifier is given same data as input, so that the final outcome from the combined classifiers is made based on the outputs of the individual classifiers obtained individually. On the other hand, in the conditional or serial architecture, individual classifiers are applied one by one in sequence, implying some specific ordering or ranking for them. When the base or primary classifier cannot be trusted to classify a given data or object for example may be because of the low confidence or support in its outcome, then the data is fed to a secondary classifier, adding classifiers in sequence. This technique is suitable when the cost of classifier exploitation is significant, so that the base classifier is computationally cheapest one and secondary classifiers have comparatively larger exploitation cost.

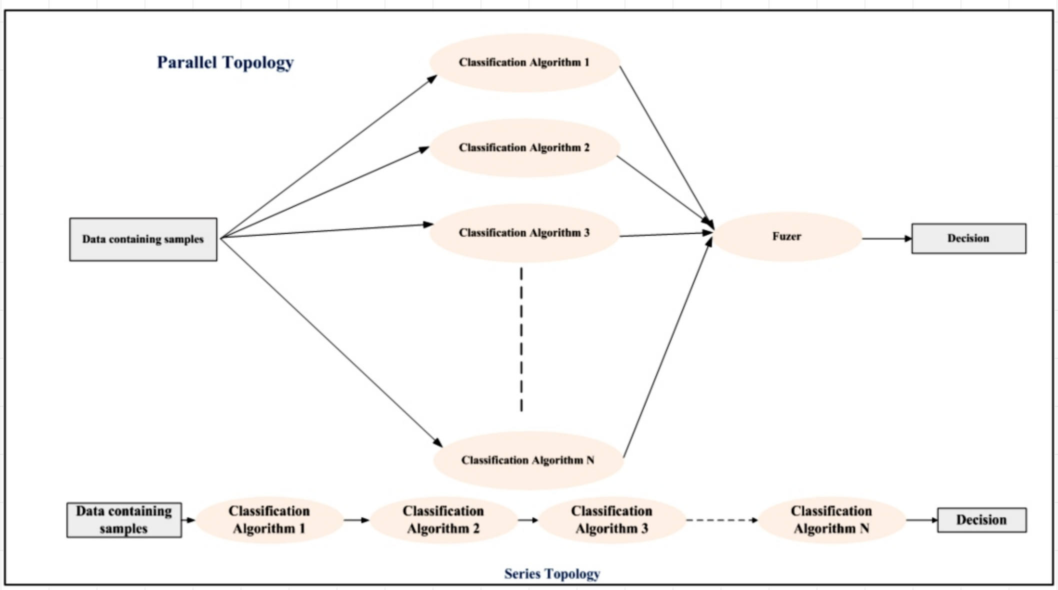

Fig. 1. Parallel & Series Topologies for Multiple / Ensemble Classifiers [xvi]

III. METHODOLOGY

This paper is extension of our previous proposed work [viii] as already mentioned in Section I. Since the experiments that are to be performed to get the EEG data will consider oscillatory rhythms, the objective of this paper is to apply different Machine learning Classifiers using Matlab on the EEG dataset that consists of EEG signals from three normal subjects during three labeled sessions consisting of repetitive self-paced right or left hand movement imagery, and the imagination of words starting with the given specific random letter.

A. Precomputed Features - PSD

Firstly, the raw EEG potentials were spatially filtered using a surface Laplacian filters. Then, at the rate of 16 times per second i.e. every 62.5 ms -- the power spectral density (PSD) in the frequency band of 8-30 Hz is evaluated over the last second of EEG data having the frequency resolution of 2 Hz for the 8 centro-parietal electrodes C3, Cz, C4, P3, Pz, CP1, CP2, and P4. The resulting EEG sample is a 96-dimensional feature vector (product of 8 electrodes and 12 frequency components). The organizers of the competition have provided the data in this pre-computed form [xxiii]. The Power Spectral Density (PSD) can be computed from the Fourier transform of the Auto correlation function of a signal. It basically elaborates how the signal squared value or the signal power is spread in terms of frequency [xxiv].

Fig. 2 shows the flowchart of each stage of the methodology out of which the steps till feature extraction have been done by the competition organizers. The stage of classification to predict classes from extracted features is performed in this work. As per the requirements of the competition, firstly the average has been calculated for each 0.5 second [xxii]. The Statistical Toolbox of Matlab is used for this work. LDA, Decision Tree and Naïve Bayes classification algorithms are applied and compared for the given dataset.

B. Linear Discriminant Analysis (LDA)

Let $X_1 = \{x_1^1, x_1^2, \ldots x_1^l\}$ and $X_2 = \{x_2^1, x_2^2, \ldots x_2^l\}$ be samples from two different classes. Linear discriminant is given by the vector *w* which maximizes

$$J(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_W \omega} \qquad (1)$$

where

$S_B := (m_1-m_2)(m_1-m_2)^T$ and
$S_w := \sum \sum (x-m_i)(x-m_i)^T$

are the between and within class scatter matrices and m_i determines the direction that maximizes the projected class means (numerator) whereas minimizing the classes variance (the denominator) [xiv]. LDA works on following three assumptions [xv]:

- Features belonging to each class are Gaussian distributed
- The Gaussian distributions for each class have same covariance matrix.
- True class distributions are already known.

In this work the LDA classifier computes the sample mean of each of the three classes. As per equation (1), the sample covariance matrix is computed by first subtracting the mean of each class from samples of that class and then the empirical covariance matrix is evaluated.

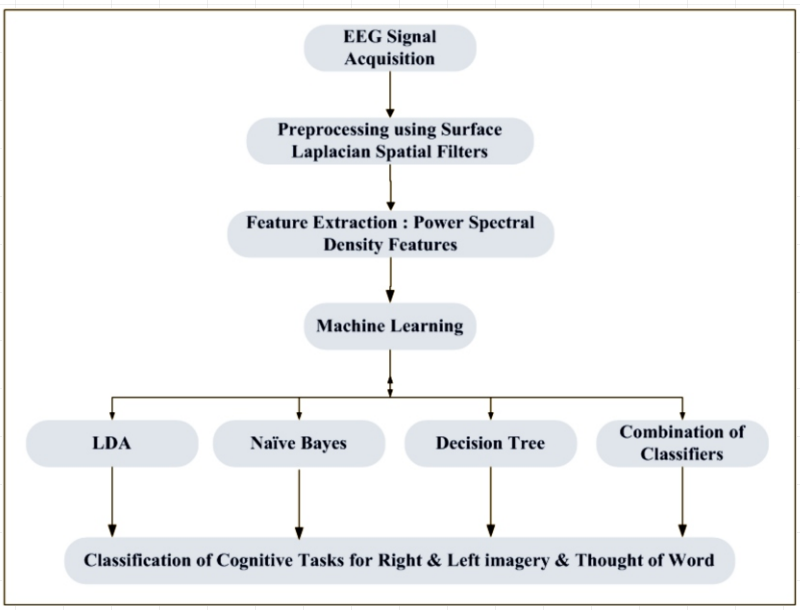

Fig. 2. Flowchart explaining the complete methodology

A. Naïve Bayes

A naive bayes classifier is one of the methods of supervised learning algorithms that is based on applying bayes theorem which says:

$$p(c_j \mid d) = \frac{p(d \mid c_j)\; p(c_j)}{p(d)}$$

where

$p(c_j|d)$ is the probability of instance d being in class c_j,
$p(d|c_j)$ is the probability of generating instance d given class c_j,
$p(c_j)$ is the probability of occurrence of class c_j, and
$p(d)$ is the probability of instance d occurring

It is a simplified Bayesian probability model which assumes that the features are independent and the probability of one attribute will not affect the probability of other attribute. It works well on classification but it states that the error occurs due to three factors including bias, variance and training data noise [xvii]

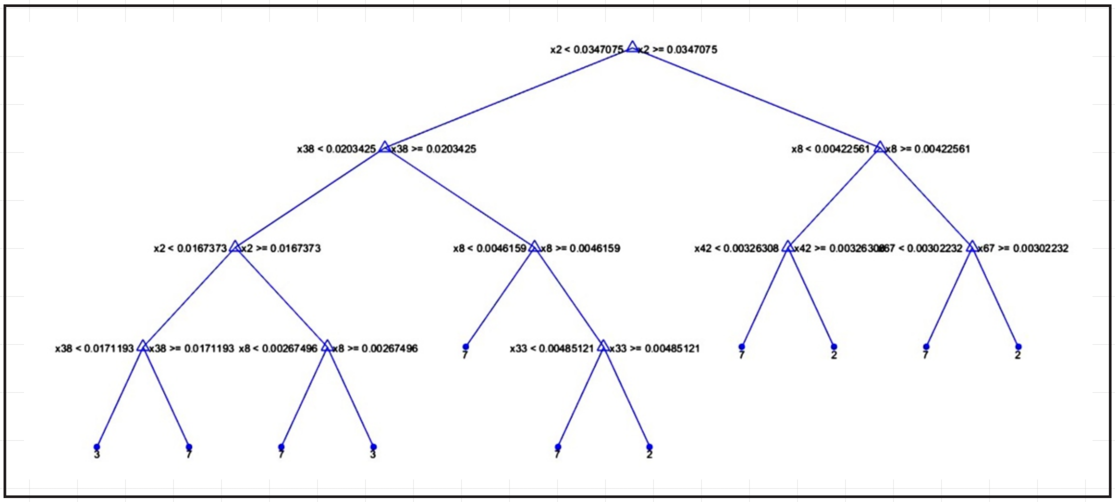

Fig. 3. Decision Rules based on initial setting for 10 splits

D. Decision Tree

Decision Tree is one of the non-parametric supervised learning algorithms used for classification and regression. A decision tree illustrates graphically the decisions to be taken, the events that may take place, and the outcomes associated with different combinations of events and decisions. Probabilities are allocated to the events, and values are evaluated for each outcome. Using this approach, complex decisions are split into simpler decisions. Structure resembling a tree is formed as a result of this classification method. [xviii]. Decision tree (DT) can lead to instability since small variations in data may form an entirely different tree that was earlier generated. This problem can be generally catered by combining decision tree classifiers using an ensemble approach.

In this work, Classification And Regression Tree (CART), one of the widely used DT algorithms is used. CART is a technique that uses tree structure in binary form having two branches at each node. The first internal node also called the root node is the point where the process of tree building initiates such that the entire data is split into two subsets. The splitting of each subset is based on a parameter for predefined criterion. According to this criterion, a test is conducted at each step to find out the most suitable feature that gives the best separation of the training samples [xxv]. This work uses the GINI index criterion to build the tree. The Gini index provides a measure of the impurity degree in a dataset. For any dataset, the Gini's Diversity Index (gdi) or Gini index of a node is computed using

$$1 - \sum_{i} p^2(i),$$

where the summation term is over the classes i at the node, whereas p(i) is the fraction of classes for class i that reach the node. A node with only one class has Gini index 0 is termed as a pure node. Other than pure node, the Gini index is always positive.

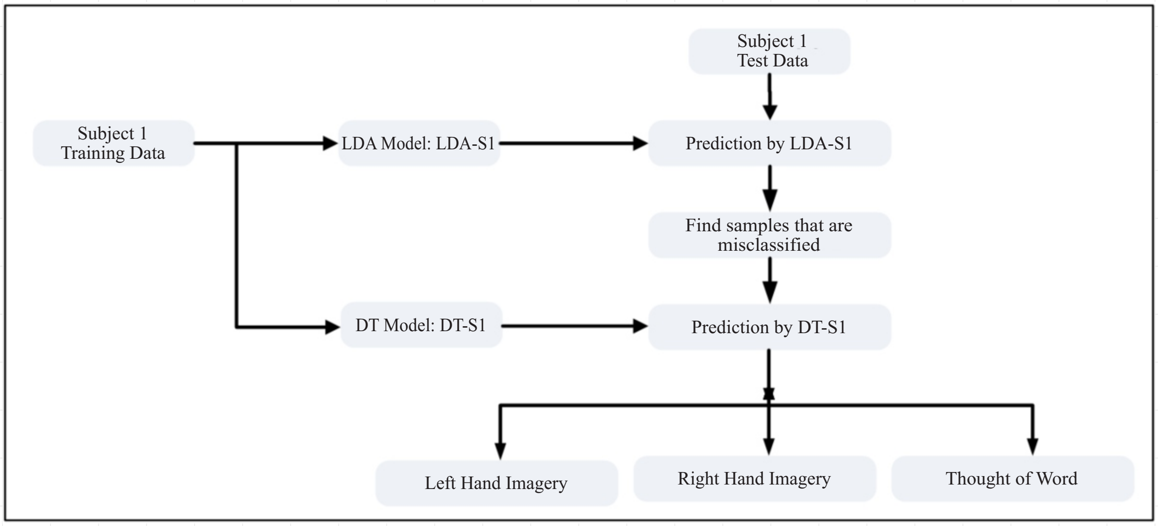

Fig. 4. Architecture for Ensemble Classifier based on LDA as base and Decision Tree as secondary classifier

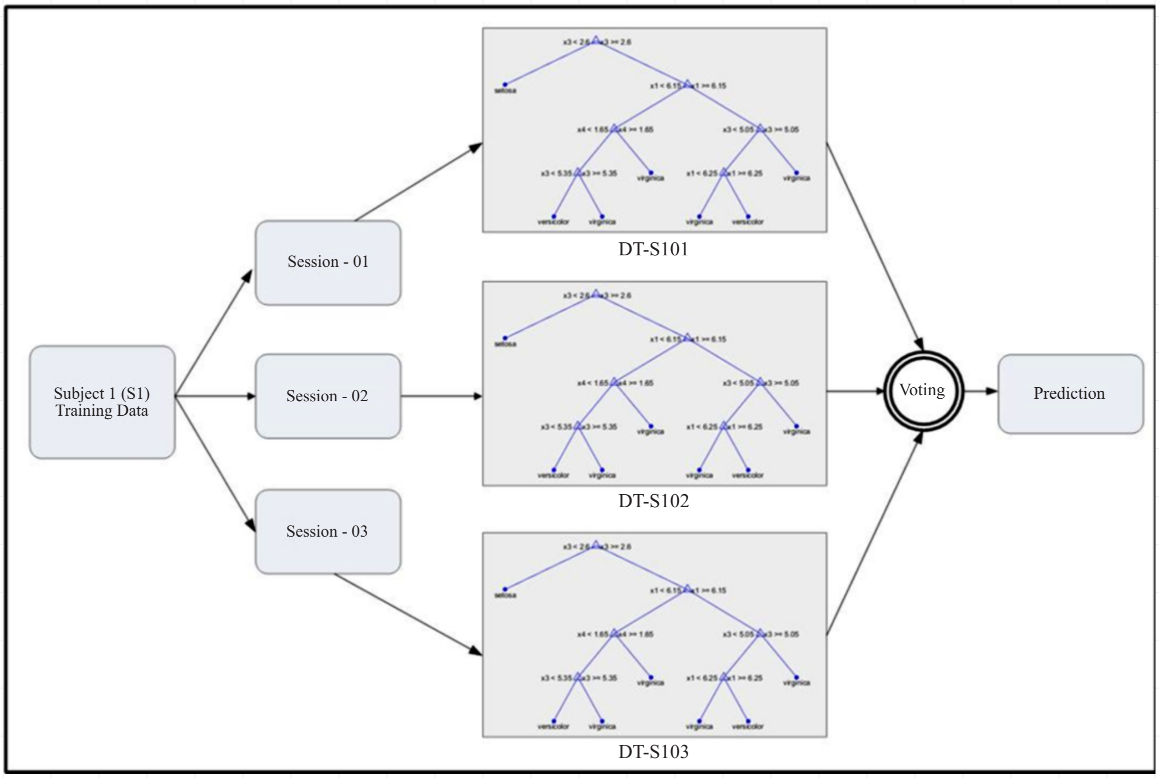

Fig. 5. Framework for Ensemble Model based on Random Forest

Followings steps are performed to generate decision tree in this work:

1. Beginning with entire input data, examine possible binary splits for each predictor
2. A split having best optimization criterion based on GINI index is selected
3. Splitting is stopped whenever any of the following is achieved: Either the node is pure or the parameter for 'Maximum Number of Splits' (MaxNumSplits) is achieved
4. Split is imposed.
5. Same procedure is repeated for the two child nodes.

In this work, firstly the MaxNumSplits is set to 10. The decision rules based on this criterion are mentioned in Fig. 3. Initially, out of 96 features, few are retained by the classifier as obvious from the figure. In order to improve the accuracy, the splits increased and decision rules are modified.

E. Multiple / Ensemble Classifier System based approach

In order to improve the performance of the classifiers, the approach to combine classifiers based on Multiple classifier system is proposed and applied as shown in Fig. 4. The approach used in [xxi] for face detection in case of various types of occluded faces has been followed. Combination of LDA and decision tree classifiers is used to combine the results from individual classifiers where LDA is used as base or primary classifier while the decision tree is integrated as secondary one. Firstly, the LDA and decision tree classifier models are trained by providing the same training set. The test data is then provided to the LDA trained classifier. The samples or instances that are misclassified by LDA are fed to the decision tree classifier.

The classifiers based on decision tree are ensembled by means of random forest (RF). A RF classifier defined as an ensemble of several decision trees where each casts a vote for a majority decision for the class selection of input data sample. Each session of a subject produced specific decision tree classifiers.

In Fig. 4, the approach of combination of classifiers for Subject 1 has been elaborated. Firstly, the training data from session 1 for Subject 1 is fed as input to the classifiers LDA and named as as LDA-S1. Then training data from session 3 of the same subject is given as test data to the trained classifiers LDA-S1 for the prediction of the class of the mental task produced by the subject. The outcome is compared with the actual class labels and those instances that are missclassified by LDA-S1 are retained only for the input to the secondary classifier based on the decision tree ensemble. Each session of subject 1 is used to train random forest named as DT-S1. Thus, DT-S1 in turn comprised of ensembles of random forests trained from other two sessions of specific subject. Fig. 5 presents the framework of proposed ensemble based on random forest. Subject I is considered where random forests are trained from other sessions. For the three sessions of subject 1, they are termed as DTS101, DTS102 and DTS103. If session 3 is considered as test data, the ensemble will consider DTS101 and DTS103 for prediction Voting is performed for final decision of the predicted class label.

TABLE II
RE-SUBSTITUTION ERROR FOR EACH TRAINING SESSION FOR THREE SUBJECTS

Subject 1:

Classifier	Training Data 01	Training Data 02	Training Data 03
LDA	.0963	.0737	.0695
NB	.2431	.2419	.2489
Decision Tree	.0619	.0323	.0426

Subject 2:

Classifier	Training Data 01	Training Data 02	Training Data 03
LDA	.1682	.1574	.1290
NB	.2788	.2477	.2788
Decision Tree	.0853	.0648	.0691

Subject 3:

Classifier	Training Data 01	Training Data 02	Training Data 03
LDA	.1963	.1495	.2023
NB	.3879	.2780	.3070
Decision Tree	.0724	.0467	.0744

IV. RESULTS

Classifiers based on LDA, NB and Decision Tree algorithms are applied on the EEG dataset described in following section.

A. Dataset

BCI Competition III, Dataset V is obtained from the BCI Competition III website [xxii] provided by the IDIAP Research Institute in Switzerland [xxiii]. It consists of EEG signals from three normal subjects during three labeled sessions (each session is 4 minutes long) without feedback. During each session, the subject performed three brain activities (in random manner) for almost 15-20 seconds duration, resulting in more or less 16 repetitions of each cognitive task that consisted of repetitive self-paced right or left hand movement imagery, and the imagination of words starting with the given specific random letter. EEG recordings were performed with using a Biosemi system having a cap with 32 EEG electrodes placed at standardized positions as per the International 10-20 system. 512 Hz is set as the sampling rate. Signals are recorded at full DC having no incorporation for artifact correction or rejection.

The classification accuracy of each model is depicted and then their respective accuracies are compared. For each of these classification procedures, the performance is measured by means of two parameters: the resubstitution error and cross validation error. The re-substitution error also know as misclassification error has been computed for each of three training session for all three subjects as shown in Table II.

TABLE III
RE-CROSS-VALIDATION ERROR FOR EACH OF THE THREE SUBJECTS

Subject 1:

Classifier	Training Data 01	Training Data 02	Training Data 03
LDA	.2271	.2166	.2197
NB	.2936	.3065	.2915
Decision Tree	.3394	.3641	.3049

Subject 2:

Classifier	Training Data 01	Training Data 02	Training Data 03
LDA	.3710	.2986	.2972
NB	.3618	.3125	.3456
Decision Tree	.4447	.3657	.3502

Subject 3:

Classifier	Training Data 01	Training Data 02	Training Data 03
LDA	.4369	.3107	.4442
NB	.5140	.3949	.4605
Decision Tree	.5397	.4673	.6000

The cross-validation errors for each of the subjects are shown in Table III. In order to validate the model, 10 fold cross-validation is performed in which the data is randomly split into ten (10) almost equal-sized partitions, and 9 out of 10 folds are used as training data while the remaining one is retained as the validation set so that in the end, every sample or instance has been used exactly once for testing purpose.

Furthermore, statistical parameters including sensitivity, specificity and confusion matrix are also used for performance evaluation of the classifiers. Specificity is defined as the measurement of fraction of negatives that are correctly identified as such (e.g., the percentage of samples of left hand movement imagination that are correctly identified as not having this state). It is also called the true negative rate. Sensitivity is defined as the measurement of fraction of positives that are correctly identified as such. For example, the percentage of samples of left hand movement imagination that are correctly identified as having this state. Sensitivity is also called true positive rate. In terms of formula, they can be written as:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

where, TP, TN, FP and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Confusion matrix and statistical measures are mentioned in Table IV and V respectively. Table VI lists the maximum accuracies obtained from each of the classifier as well as ensemble classification algorithm for each subject. The accuracies obtained by the winner of the BCI competition III dataset V are also mentioned [xxii].

V. DISCUSSION

From Table II, the least resubstitution error is obtained in case of Decision tree algorithm while LDA has lesser to that of Naïve Bayes. So, it apparently seems that DT has produced the best results but if the comparison is performed using cross validation error, it is found that here LDA outclasses DT. The cross-validation errors for each of the subjects are shown in Table III. Due to difference between resubstitution and cross-validation errors of DT, it can be inferred that the decision tree has overfitted the data. It seems that the tree classifies the original training data in good manner, but design of the tree is little sensitive to the specific training set that's why its performance on any new test dataset may likely to debase.

While comparing the results from each of the single classifiers, it is observed that Naïve Bayes has performed best for the given dataset as shown in Table II. For subjects 1 and 2, the Naïve Bayes classifier produces highest accuracy while for subject 3 the maximum accuracy is achieved from LDA classifier. From Table II, it is also cleared that the results does not depend only the classifier, but also on the training session of given subject. If the user is feeling lethargic during that session, the performance can degrade significantly.

TABLE IV
CONFUSION MATRICES OF THE CLASSIFIERS FOR SUBJECT 1 SESSION 3
(LABELS 2, 3 AND 7 SHOW LEFT HAND, RIGHT HAND AND RANDOM WORD IMAGINATIONS RESPECTIVELY)

Classifiers	**True Labels**	**Predicted Labels** 2	3	7
LDA	**2**	89	13	9
	3	15	85	34
	7	38	24	139
DT	**2**	79	16	11
	3	31	81	27
	7	32	25	144
NB	**2**	109	30	15
	3	15	77	20
	7	18	15	147
Ensemble	**2**	106	24	12
	3	18	101	3
	7	8	19	155

A. Statistical Measures

The confusion matrices presenting the results of the classifiers applied are given in Table IV. Labels 2, 3 and 7 show left hand, right hand and random word imaginations respectively. From these matrices, we can tell the frequency by which an EEG sample is misclassified as another. To find out, how well each class is identified by the classifier, we need to look at Table V for sensitivity and specificity values. From Table V, it is obvious that samples for random word generation are more identifiable as compared to other instances. Sensitivity for each classifier in case of class 7 achieve highest values. And for the proposed classifier attains the highest value 91.18%. The Naïve Bayes classifier identifies class 2 labels more efficiently as compared to class 3 samples. Similarly, the proposed approach outperforms other classifiers for identification of label 3 and achieved highest sensitivity of 80.33%.

TABLE V
STATISTICAL MEASURES OF THE CLASSIFIERS

Classifiers	True Labels	Statistical Parameters (%)		
		Sensitivity	Specificity	Classification Accuracy
LDA	2	62.68	91.06	70.18
	3	69.67	82.31	
	7	76.37	73.73	
DT	2	55.63	89.29	68.16
	3	66.39	79.36	
	7	79.12	73.73	
NB	2	76.76	83.27	74.66
	3	63.11	87.97	
	7	80.77	84.93	
Ensemble	2	80.30	87.67	81.16
	3	70.14	92.55	
	7	91.18	88.46	

B. Comparison with State of the Art

Table VI lists the maximum accuracies obtained from each of the classifier for each subject. Here we will only talk about the maximum accuracy obtained for a particular subject across different training samples. In Table VII, the accuracies obtained by the winner and 1st runner up of the BCI competition III dataset V, are also mentioned to compare our results with them [xxiii]. Here, it is observed that by combining the classifiers for the same training data, the accuracies are improved significantly and can be compared with the one obtained by the winner of the competition. The winner works with the precomputed samples. It has been trained off-line with data of first three sessions and it is structured in three stages: preprocessing and feature extraction statistical discrimination, and online discrimination improvement. First of all, before any analysis data are transformed by means normalizing of each PSD sample. Each spectral component of channel i from sample t is normalized dividing by the energy of PSDt(i). With data normalized, the feature extraction process is guided by canonical variates transform (CVT), a generalization of Fisher's linear discriminant function to more than two groups. This transformation permits the projection of a p-dimensional dataset X to be classified into c classes in a (c-1)-dimensional feature space where classes separation is maximized [xxii-xxiii].

Results of the first runner up are also mentioned in Table VII. In this work the r^2 values are plotted to select good discriminative component of the input data as input feature for the purpose of classification. The features are normalized to an interval [0,1]. Support Vector Machine (SVM) classifier is obtained from the training data and five fold cross-validation is applied [xxii-xxiii].

In our approach, firstly LDA as a weak classifier performs classification. By introducing randomness in building random forests, the classifiers become robust and cause them to have a good performance when the data have many outliers which is our case. Predictions from random forests are considered on the basis of majority voting approach resulting in significant increase in classification accuracy. Results mentioned in Table VII are plotted in Fig. 6.

TABLE VI
MAXIMUM CLASSIFICATION ACCURACIES

Classifier	Sub. 1	Sub. 2	Sub. 3
• LDA	70.17	55.07	51.16
• NB	74.66	60.13	46.49
• Decision Tree	68.16	56.68	46.26
• Ensemble - LDA combined with DT	81.16	70.27	56.51

TABLE VII
COMPARISON WITH STATE OF THE ART

Classifier	Sub. 1	Sub. 2	Sub. 3
Author's Results based on Ensemble - LDA combined with DT	81.16	70.27	56.51
Winner Result based on CVT [xxii-xxiii]	79.60	70.31	56.02
1st Runner Up result based on SVM [xxii-xxiii]	78.08	71.66	55.73

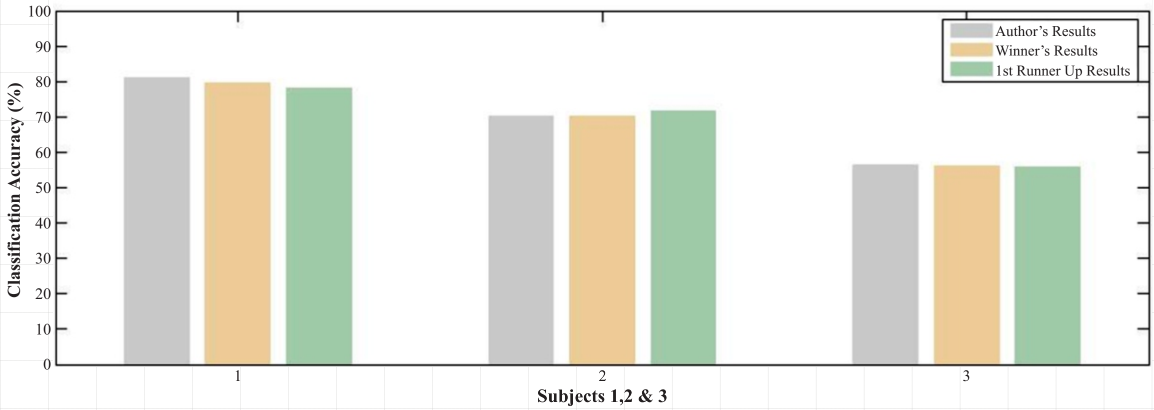

Fig. 6. Results from Table VII are plotted as bar chart

VI. Conclusion & Future Work

The main objective of this study is to demonstrate the application and comparison of machine language techniques using LDA, Naïve Bayes and Decision Tree to predict the cognitive state of the subject. The dataset V from BCI Competition III is used for this purpose. The features based on Power Spectral Density have been given as input to the classifiers. In order to further improve the accuracy, combination of classifier is used where LDA is used as the primary classifier while decision tree from random forests as the secondary one. The results are compared with that of the winner of the competition and it is found that the classification accuracies obtained after the combination of classifier has improved significantly. While comparing the results from each of the single classifiers, Naïve Bayes has performed best. Two parameters are used to assess the performance of the classifier namely resubstitution error and cross-validation error. For statistical analysis, sensitivity and specificity for each classifiers are evaluated.

In future these machine learning algorithms can be applied for the classification stage of our proposed work [viii]. And it will be evaluated if the results are obtained accordingly as we have found in this work using BCI Competition dataset.

Acknowledgments. This work is funded by Higher Education Commission (HEC), Pakistan and is being conducted and supervised under the 'Intelligent Systems and Robotics' research group at Computer Science (CS) Department, Bahria University, Karachi, Pakistan..

References

[I] F. Lotte, "Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications," INSA de Rennes, 2008.

[ii] B. Blankertz, M. Krauledat, G. Dornhege, J. Williamson, R. Murray-Smith, and K.-R. Müller, "A note on brain actuated spelling with the Berlin brain-computer interface," in *Universal Access in Human-Computer Interaction. Ambient Interaction*, ed: Springer, 2007, pp. 759-768.

[iii] J. Höhne, M. Schreuder, B. Blankertz, and M. Tangermann, "A novel 9-class auditory ERP paradigm driving a predictive text entry system," *Frontiers in neuroscience,* vol. 5, p. 99, 2011.

[iv] M. C. F. Castro, A. Araujo Masiero, F. Theoto Rocha, and P. T. Aquino, "Motor imagery recognition and its cerebral mapping," in *Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC), 5th ISSNIP-IEEE*, 2014, pp. 1-5.

[v] R. Chai, S. H. Ling, G. P. Hunter, and H. T. Nguyen, "Mental non-motor imagery tasks classifications of brain computer interface for wheelchair commands using genetic algorithm-based neural network," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1-7.

[vi] M. Naeem, C. Brunner, and G. Pfurtscheller, "Dimensionality reduction and channel selection of motor imagery electroencephalographic data," *Computational intelligence and neuroscience,* vol. 2009, 2009.

[vii] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage,* vol. 56, pp. 387-399, 2011.

[viii] N. Irtiza and H. Farooq, "The study of fear-induced power modulations for Cognitive Man-Machine Communication," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on*, 2015, pp. 346-351.

[ix] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing,* vol. 129, pp. 94-106, 2014.

[x] G. G. Knyazev, J. Y. Slobodskoj-Plusnin, A. V. Bocharov, and L. V. Pylkova, "Cortical oscillatory dynamics in a social interaction model," *Behavioural brain research,* vol. 241, pp. 70-79, 2013.

[xi] S. Dähne, F. C. Meinecke, S. Haufe, J. Höhne, M. Tangermann, K.-R. Müller, and V. V. Nikulin, "SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters," *NeuroImage,* vol. 86, pp. 111-122, 2014.

[xii] P. Chum, S.-M. Park, K.-E. Ko, and K.-B. Sim, "Optimal EEG feature selection by genetic algorithm for classification of imagination of hand movement," in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*, 2012, pp. 1561-1566.

[xiii] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 4, p. R1, 2007.

[xiv] B. Scholkopft and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX,* vol. 1, p. 1, 1999.

[xv] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—a tutorial," *NeuroImage,* vol. 56, pp. 814-825, 2011.

[xvi] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion,* vol. 16, pp. 3-17, 2014.

[xvii] G. Keerthika and D. S. Priya, "Feature Subset Evaluation and Classification using Naive Bayes Classifier," *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org,* vol. 1, 2015.

[xviii] U. R. Acharya, R. Yanti, J. W. Zheng, M. M. R. Krishnan, J. H. TAN, R. J. Martis, and C. M. Lim, "Automated diagnosis of epilepsy using CWT, HOS and texture parameters," *International journal of neural systems,* vol. 23, p. 1350009, 2013.

[xix] J. He, H.-J. Hu, R. Harrison, P. C. Tai, and Y. Pan, "Transmembrane segments prediction and understanding using support vector machine and decision tree," *Expert Systems with Applications,* vol. 30, pp. 64-72, 2006.

[xx] S. Sun, C. Zhang, and D. Zhang, "An experimental evaluation of ensemble methods for EEG signal classification," *Pattern Recognition Letters,* vol. 28, pp. 2157-2163, 2007.

[xxi] K. Ichikawa, T. Mita, O. Hori, and T. Kobayashi, "Component-based face detection method for various types of occluded faces," in *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, 2008, pp. 538-543.

[xxii] B. C. III. (2004/2005). *BCI Competition III 2004/2005*. Available:http://www.bbci.de/competition/iii/results/#martigny

[xxiii] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. R. Millan, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 14, pp. 153-159, 2006.

[xxiv] A. B. Benevides, T. F. Bastos Filho, and M. Sarcinelli Filho, "Pseudo-Online Classification of Mental Tasks Using Kullback-Leibler Symmetric Divergence," *Journal of Medical and Biological Engineering,* vol. 32, pp. 411-416, 2012.

[xxv] W. Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 1, pp. 14-23, 2011.