Refernce shob ek spacing diye kor, pura bisri lage chokhe

Motor Imagery EEG Signal Classification Using Random Subspace Ensemble Method

Zabir Al Nazi, Md. Monirul Islam, A.B.M. Aowlad Hossain
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology (KUET) Khulna-9203, Bangladesh.
zabiralnazi@codeassign.com,mohammad.monirul.islam.10@gmail.com,aowlad0403@gmail.com

***Abstract*—Classification of EEG signals for Brain Computer Interface (BCI) has great impact on people having various kinds of physical disabilities~~ who are 'locked-in'~~. Motor imagenry (MI) EEG signals of hand and leg movement classification can help people whose limbs are replaced by prosthetics. In this paper, Random Subspace Ensemble (RSE) method was proposed for improving prediction accuracy. The method was demonstrated on four different subjects and a hybrid dataset of two subjects' data combined. Principal Component Analysis (PCA) was introduced for dimensionality reduction of the feature space. A comparative analysis was studied where Random Subspace Ensemble method outperformed other machine learning models. Furthermore, the model showed better performance with reduced feature set generated by PCA. The maximum accuracy obtained was 95.8% with original feature dimension and 87.5% with PCA features. The finding of the study will ~~help the people who has lost their limbs~~** contribute to the brain computer interface research.

for motor imagery EEG classification.

classification

***Index Terms*—Electroencephalography (EEG), Brain Computer Interface (BCI), Principal Component Analysis (PCA) and Random Subspace Ensemble (RSE) Method.**

I. Introduction

A Brain Computer Interface (BCI) is the bridge between human brain and computer. It has given us the potential to look into the human brain and study the regions which was previously impossible. By analyzing electroencephalogram (EEG), we can study human brain and with the aid of BCI we can make important decisions about how our brain works to communicate and control[1]. Basically there are two kinds of BCI systems that control the exoskeleton, i.e. BCI based on Motor Imagery (MI) and BCI based on Steady State Visual Evoked Potentials (SSVEPs)[2]. The main advantage of SSVEP based control is that it requires less time to train but the limitation of this system is that it results in a higher false detection. Another drawback is that it aims on the communication with environment neglecting functional rehabilitations made by patients [3].

As computational power of computers have increased researchers have taken numerous efforts to study the EEG signal like classifying MI classes (imagined movements of left and right hand). Previous efforts were concentrated on categorizing the MI EEG signals into phenomena of Event Related Desynchronization (ERD) and Event Related synchronization (ERS)[4]. On the other hand, band power, interval variance, auto regressive model and spectral decomposition, temporal spectral evolution task related power increase and decrease etc [5] are included in quantification measurements of ERD/ERS[6][7].

Time, frequency and wavelet all three domains of EEG signals were analyzed during the motor imagery EEG classification[8]. Extracted features from these domains were analyzed with Linear Discriminant Analysis (LDA) algorithm to obtain the optimal features. Then these features were used for ANN classifier which resulted 83.6% accuracy. Another interesting approach was made by the authors in[9], who tried to classify imagined words with conventional EEG classification and sonification and textification. While textification outperformed conventional and sonification approaches by average accuracy of 83.34%.

In this paper, we will decompose EEG signal using DWT and de-noise the signal. From selected sub band, we will extract features and we will use PCA to reduce the dimensionality of the features. Then RSE method will be used to classify the signals. and finally we will use PCA ..

and comparison will be made with other classification models.

Rest of paper is organized as : in section II, Proposed methodology is discussed. In next section detailed analysis of result was given and finally there is the conclusion of this work.

organized as follows: a conclusion of this work is drawn.

II. Methodology

The overall methodoloy used in this paper is given in the fig.1. At first, the signal was preprocessed using DWT and noise was removed. Then initially 50 features (and for one single case it was 60) were selected. PCA was also used to reduce dimensionality of the features. After that, these features were used to classify the EEG signal. For classfication, RSE algorithm was used. The overview of proposed method is : feature extraction for raw EEG signal and dimension reduction of feature and this features are then used for classification using RSE method finally result was evaluated.

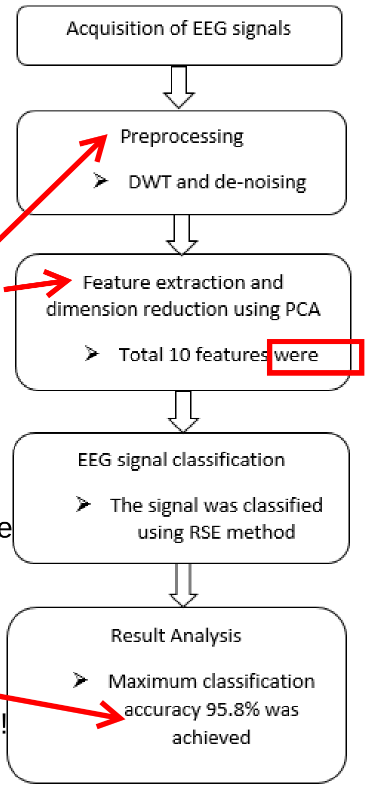
A. Denoising EEG signal

For noise removal of biomedical signals which are easily contaminated with artifacts, wavelets are utilized with shapes similar to the corresponding signal class. In this scheme, by applying DWT, the signal is decomposed into its wavelet coefficients. By using a threshold discrimination filter, coefficients contributing to noise constituents are separated and removed after selecting a predefined threshold value. For this work,

DWT
Subband Selection
Denoising

Font bigger

10 features



Random Subspace Ensemble method

WTF alignment?!

Figure 1. Proposed methodology

orthogonal Coiflets 2 (coif2) wavelets were used in DWT to remove noise from EEG signals. [10]
DWT provides multi-resolution analysis of signal using basis

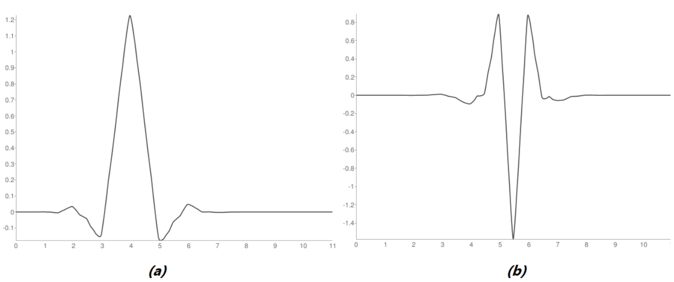


Figure 2. (a) Scaling function ϕ and (b) Wavelet function ψ

function known as wavelet function. It is defined as:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_o^j}}\psi\left(\frac{t - k\tau_o s_o^j}{s_o^j}\right) \qquad (1)$$

Mallat[11] introduced quadrature mirror filters for an efficient way of implementation by passing the signal through a series of low-pass (LP) and high-pass (HP) filter pairs. The outputs from the low and high pass filters are referred to as approximation (A_1) and detail (D_1) coefficients of first level respectively. This process is continued for more decomposition levels. The coefficients represent frequency contents of the original signal within different bands.

The detailed coefficients d_k^l are adapted with a threshold coefficient filter and a threshold value. In wavelet denoising method, the absolute value of detailed coefficients is considered, coefficients smaller than the threshold are zeroed out. Only large coefficients are considered to contribute to the useful information of the original EEG signal. In this work, a shrinkage function based soft thresholding is used because hard thresholding generates artifacts due to discontinuity[17]. The soft thresholding rule is defined by:

$$d_k^l \begin{cases} 0, & |d_k^l| < \lambda \\ sng(d_k^l)(d_k^l) & |d_k^l| \geq \lambda \end{cases}$$

where $sgn(d_k^l)$ is mathematical sign function.

B. Features selection & dimension reduction with PCA

Features from both time and frequency domain can be used for EEG classification. But random features can decrease the performance of classification. In [12], authors proposed features for classification of signals in both time and frequency domain. 10 features are selected from them according to those features' performance.

I. Average amplitude change (AAC) is formulated as

$$ACC = \frac{1}{N}\sum_{i=1}^{N-1}|x_{i+1} - x_i|$$

II. Difference Absolute Standard Deviation Value (DASDV) is like RMS feature, can be defined by

$$DASVD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1} - x_i)^2}$$

III. Mean Absolute Value(MAV) or first order of v-Order features (V1), average rectified value (ARV), integral of absolute value (IAV), averaged absolute value (AAV) is defined as

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

IV. Modified Mean Absolute Value Type 2 (MAV2) is an expansion of MAV feature. For improving smoothness it is used. MAV is defined as

$$MAV = \frac{1}{N}\sum_{i=1}^{N}w_i|x_i|$$

$$w_i = \begin{cases} 1, & if \quad 0.25N \leq i \leq 0.75N \\ \frac{4i}{N}, & else \ \ if \quad i < 0.25N \\ \frac{4(i-N)}{N}, & otherwise \end{cases}$$

V. Myopulse Percentage Rate (MYOP) is calculated as

$$MYOP = \frac{1}{N}\sum_{i=1}^{N}[f(x_i)]$$

$$f(x) = \begin{cases} 1, if \quad x \geq threshold \\ 0, otherwise \end{cases}$$

Here, threshold is pre-defined.
VI. Root Mean Square (RMS) is defined as following,

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$$

VII. 2^{nd} Spectral Moments(SM2) is a statistical approach to extract power spectrum of EEG singal and it is defined as

$$SM2 = \sum_{i=1}^{M} P_i f_i^2$$

VIII. Log Detector (LOG) is defined as

$$LOG = \exp(\frac{1}{N}\sum_{i=1}^{N} log(|x_i|))$$

IX. Waveform Length (WL) is used to measure the complexity of EEG signal and is defined as

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i|$$

X. Integrated EEG is defined as sum of absolute value of amplitude.

$$IEEG = \sum_{i=1}^{N} |x_i|$$

Principal component analysis (PCA) is a renowned scheme for feature extraction. PCA linearly transforms a high-dimensional feature vector into a low-dimensional one where components are uncorrelated, by calculating the eigenvectors of the covariance matrix of the original inputs.
Let, for centered input vectors $x_t(t = 1, ..., l$ and $\sum_{t=1}^{l} x_t = 0)$ and each of x_t has m dimensions i.e., $x_t = (x_t(1), x_t(2), ..., x_t(m))^T$ where $m < l$
By using PCA, we get s_t,

$$s_t = U^T x_t \tag{2}$$

where U is an m-square orthogonal matrix and it's i^{th} column u_i is the i^{th} eigenvector having a simple covariance matrix $C = \frac{1}{l}\sum_{t=1}^{l} x_t x_t^T$. That means PCA solves the eigenvector problem (3) first.

$$\lambda_i u_i = C u_i, where, i = 1, ..., m \tag{3}$$

where λ_i and u_i are one of eigenvaules of C and eigenvectors respectively. So, orthogonal transformation x_t,

$$s_t(i) = u_i^T x_t, where, i = 1, ..., m \tag{4}$$

And these components are called principal components. Using only first few eigenvectors in descending order reduces the number of principal components. Properties of PCA are given below:
1) Dimension reducibility.
2) $s_t(i)$ have sequential maximum variance and all of them are uncorrelated.
3) Mean squared approximation of first several principal components are minimal [13].

In our proposed method, for each of the C channels K features were chosen, after concatenating features for all channels, a feature vector was generated with dimension $C * K$. With PCA 10 principal components were extracted which resulted in a sparse feature set with only 10 features. The low dimensional feature vector was employed for further classification.

C. Classification Method

To improve prediction accuracy of the classifier (Discriminant Analysis) RSE is introduced by Ho[14]. This method works on a subspace randomly selected from the original feature space. The problem of dimensionality is optimized in this reduced space as number of subject per feature grows and useful features are retrieved [15]
RSE modifies the training dataset by sampling features, this modified dataset is used to build classifier, majority voting technique is utilized to reach the final decision.
Let, each training data point $X_i = (x_{i1}, x_{i2}, ..., x_{ik})$ in the data set $X = (X_1, X_2, ..., X_k)$ be a vector of dimension k for k features, where $i = 1, 2, 3, ..., n$. In RSE, randomly $\tilde{k}$ features are selected $(\tilde{k} < k)$, thus $\tilde{k}$ dimensional random subspace is obtained from the k dimensional feature vector.
Thus, the original dataset is modified to: $\tilde{X}^t = (\tilde{X}_1^t, \tilde{X}_2^t, ..., \tilde{X}_n^t)$, where each data point now contains $\tilde{k}$ dimensional training objects $\tilde{X}_i^t = (\tilde{x}_{i1}^t, \tilde{x}_{i2}^t, ..., \tilde{x}_{ik}^t)$ where $i = 1, 2, ..., n$ and $\tilde{k}$ features $x_{ij}^{\tilde{k}}(j = 1, 2, ..., \tilde{k})$ are selected randomly. The classifier is built upon the new random subspace $\tilde{X}^t$ and aggregation is performed t times to reach the final prediction accuracy.

Algorithm : Random Subspace Ensemble Method

I **INPUT**: training set X,label set y subspace dimension $\tilde{k}$, no. of selection t
II **for** $t = 1, 2, ..., T$
   a) Generate $\tilde{k}$ dimensional random subspace $\tilde{X}^t$ from k dimension feature space X
   b) Build a discriminant classifier $C^t(x)$
   c) Calculate weights $c_t = \frac{1}{2}log(\frac{1-e_t}{e_t})$ where $e_t = \frac{1}{n}\sum_{i=1}^{n} w_i^t v_i^t$ and

$$v_i^t = \begin{cases} 1, for\ wrong\ class\ prediction\ of X_i \\ 0, otherwise \end{cases}$$

III merge classifier $C^t(x), t = 1, 2, ..., T$ using weighted majority votes. Use weights c_t for deision rule: $\gamma(x) = \underset{y}{\text{argmax}} \sum_t f(sgn(C^t(x)), y)$ where

$$v(i,j) = \begin{cases} 1, if\ i = j; \\ 0, otherwise \end{cases}$$

f(i,j) hobe

and $\boldsymbol{y}$ is class label.

The RSE is benefited from combining the classifier and using random subspace for feature generation. In case of training data points being relatively small compared to feature dimension, the problem of small sample size is solved by generating classifier on random subspace. The model also performs better on datasets with redundant and noisy feature[16].

III. RESULT ANALYSIS

Dataset from Dr. Cichocki's Lab (Lab. for Advanced Brain Signal Processing) was used [17]. The cue based data recording paradigm consisted of MI tasks, specifically the imagination of movement of the left hand (LH), right hand (RH) and both feet (F). In this dataset, g.tec (g.USBamp) and Neuroscan (SynAmps RT) were used for recording the EEG signals. Band pass filter was used with low and high cut-off frequency of 2Hz and 30Hz respectively with sampling rate of 256Hz with a notch filter at 50Hz for g.tec and for Neuroscan device bandpass filter between 0.1Hz and 100Hz with sample rate of 250Hz was used. The signals were measured in $\boldsymbol{\mu V}$ and $\boldsymbol{V}$ for Neuroscan and g.tec respectively.

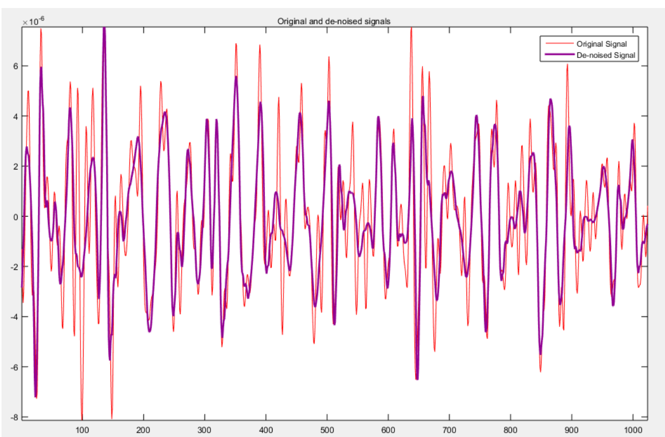


Figure 3. Original and de-noised EEG signal

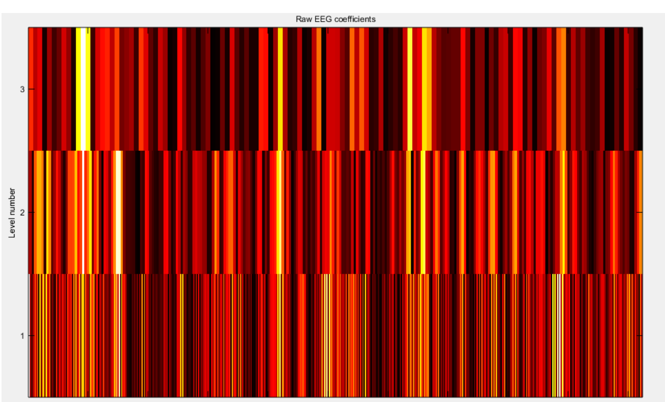


Figure 4. Heat map coefficients for EEG signal after DWT

Heat map of raw EEG coefficients after DWT

Table I
CLASSIFICATION RESULT

Classification model	Subject A	Subject C	Hybrid dataset	Subject E	Subject G
Decision Tree	71.9	80.0	71.8	77.1	62.5
SVM	84.1	85.0	79.3	85.4	70.8
KNN	79.6	80.6	79.8	89.6	67.5
Ensemble	81.9	86.7	80.7	95.8	69.2

For subjects A, C and E, data was recorded with g.tec (g.USBamp) for 4 seconds with 5 channels C3, Cp3, C4, Cp4, Cz. So, each sample contains 1024 temporal points for all 5 channels. For subject G, EEG recordings of 4 seconds with 6 channels were used.

The application of proposed Random Subspace Discriminant Ensemble method on EEG data was demonstrated on dataset SubA_5chan_3LRF, SubC_5chan_3LRF, SubE_5chan_2LR and SubG_6chan_2LR from Dr. Cichocki's Lab (cued motor imagery data with three classes: right hand, left hands and foot from 4 subjects) and on a hybrid dataset by combining data of subjects A and C. The signals were denoised with DWT with coif2 wavelets by decomposing into level 3. For each of the 5 channels 10 features were calculated, in total a feature vector of length 50 was used for classification and for 6 channels 60 features were generated. A comparative performance analysis was investigated between conventional machine learning models such as Decision Tree, SVM with linear kernel, weighted KNN and proposed Random Subspace Ensemble (RSE) method for MI classification.

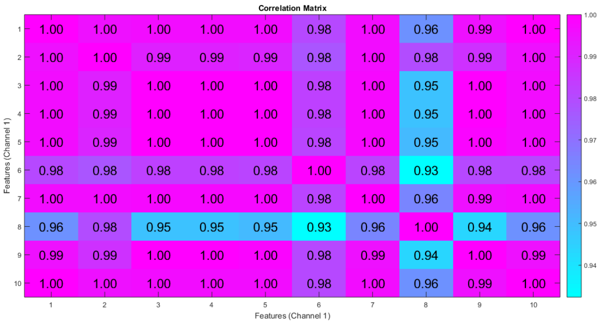
Table II
CLASSIFICATION RESULT WITH PCA

Classification model	Subject A	Subject C	Hybrid dataset	Subject E	Subject G
		Subject G			
Decision Tree	66.3	80.6	69.1	85.4	68.8
SVM	80.0	81.7	80.4	85.4	65.0
KNN	71.9	84.4	74.2	75.0	63.3
Ensemble	81.9	85.6	81.1	87.5	69.2

Ensemble method outperformed all other methods for subject C (86.7%), subject E (95.8%) and hybrid set (dataset A & C combined) (80.7%), for subject A and G Ensemble reached very close to maximum performance achieved by SVM (subject A: 81.9% and 84.1%, subject E: 69.2% and 70.8%). The results are given in table I. Then to reduce the number of features PCA was implemented. After PCA, we got 10 features previously which was 50 (for subject A, C, E) and 60 (for subject G). Performance after PCA is shown in table II.

From table I and table II we can summarize the observations as:

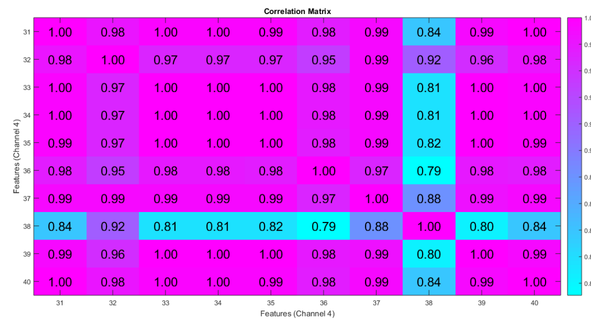
I for subject A, all the models performed similarly before dimensionality reduction but after applying PCA all the



(a)

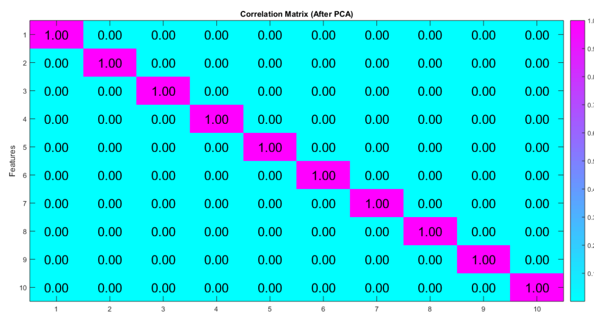
(b)

(c)



(d)

(e)



(f)

Figure 5. (a), (b), (c),(d), (e) are the Pearson correlation matrix for channel 1 to 5 respectively and (f) is the correlation matrix after applying PCA.

(a-e) Pearson correlation matrix for features of channel 1 to 5 (f) Correlation matrix after applying PCA

models showed degraded performance except proposed ensemble method which was able to generalize with small feature space.

II for subject C, after PCA, KNN's performance improved, Decision Tree and Ensembles' performance remained almost same. SVM's performance declined.

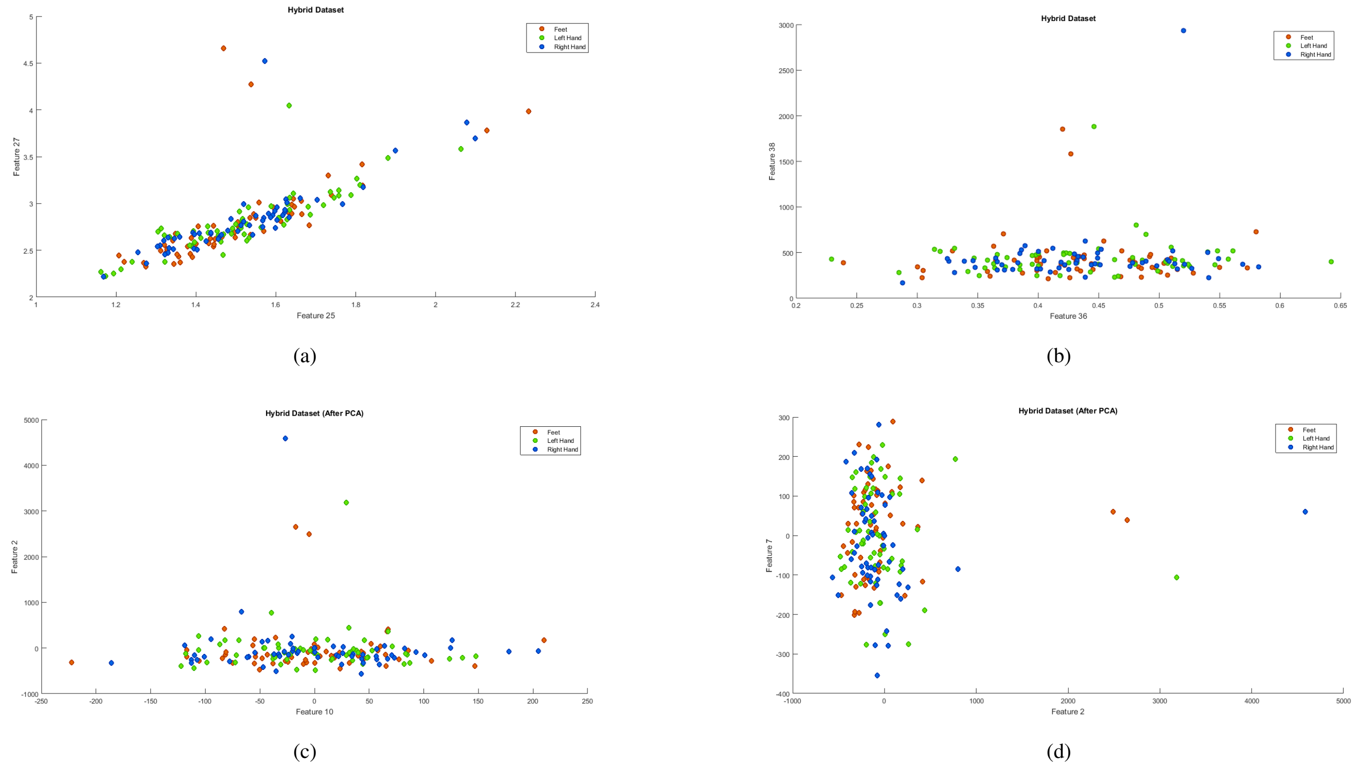
III for hybrid dataset, SVM and Ensembles' performance improved but prediction accuracy of other two models declined.

IV for subject E, performance of Decision Tree improved with reduced feature set, whereas SVMs' performance remained consistent but Ensemble method showed best performance for both cases.

V for subject G, Decision Tree showed better performance after PCA, results of SVM and KNN dropped and Ensembles' performance remained same.

Therefore, after PCA, we can say that all the algorithms showed mixed performance. But there is an important factor that must be evaluated which is the number of features. As mentioned earlier, initially there are 50 features (for subject A, C, E) and 60 features (for subject G) whereas after dimensionality reduction with PCA, there are only 10. So number of features had an impact on performance of the algorithms. For Ensemble method, another observation is that RSE method's performance does not vary that much before and after applying PCA compared to others. In case of relatively small training objects compared with number of features, RSE solved the small sample size problem by constructing classifiers in random subspace with lower feature dimension. So, when the data contains many redundant features this method performs

From fig. 5, the features are highly correlated. The less correlated features are 36 and 38. The scatter plot for these features is shown in fig. 6. After applying PCA, the features are not correlated which can be observed in the correlation matrix.



(a) (b) (c) (d)

Figure 6. (a) and (b) are scatter plot of feature 27 vs 25 and 38 vs 36 respectfully. (c) and (d) are scatter plot of feature 2 vs 10 and feature 7 vs 2 respectfully. Here (a) and (b) is scatter plot before applying PCA and (c) and (d) are scatter plot after PCA

(a) Scatter plot for features 25 and 27 (b) Scatter plot for features 36 and 38 (c) Scatter plot for PCA features 10 and 2 (d) Scatter plot for PCA features 2 and 7

better in random subspace than in original feature space. The combined decision of the weak classifiers outperforms single classifiers built on the original training dataset with complete feature space. Hence, the RSE method was able to generalize even with smaller feature set after dimensionality reduction with PCA.

IV. Conclusion

In this paper, we proposed Random Subspace Ensemble method for classifying MI EEG signals and found that this method performs better in classification task both with high dimensional feature set and reduced dimension than commonly used models such as Decision Tree, SVM and KNN. The main contribution of this work is that, we tried to find a method which generalizes with different feature dimension. Selection of important features is very critical for EEG signal classification because it helps to encapsulate the useful information embedded in the signals. The prediction accuracy of Ensemble method didn't vary much for different feature sets and showed better performance than other methods even though the training time taken was moderately high. We intend to extend our work in future by finding more suitable features that can reduce the number of features used before and using other methods of dimensionality reduction such as kernel PCA, Linear Discriminant Analysis (LDA), Generalized Discriminant Analysis (GDA) keeping in mind the improvement of prediction accuracy of the classification which will be an important aspect of real time EEG classification for Brain Computer Interface.

Acknowledgment

The authors would like to thank the authority of Khulna University of Engineering & Technology (KUET), Khulna—9203, Bangladesh for their support.

References

[1] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, Current trends in Graz brain-computer interface (BCI) research, IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 2, pp. 216219, 2000.

[2] R. Ortner, B. Z. Allison, G. Korisek, H. Gaggl, and G. Pfurtscheller, An SSVEP BCI to control a hand orthosis for persons with tetraplegia, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 19, no. 1, pp. 15, 2011.

[3] G. Pfurtscheller, C. Guger, G. Mller, G. Krausz, and C. Neuper, Brain oscillations control hand orthosis in a tetraplegic, Neuroscience Letters, vol. 292, no. 3, pp. 211214, 2000.

[4] G. Pfurtscheller and C. Neuper, Motor imagery and direct brain-computer communication, Proceedings of the IEEE, vol. 89, no. 7, pp. 11231134, 2001.

[5] R. Salmelin, M. Hmalinen, M. Kajola, and R. Hari, Functional segregation of movement-related rhythmic activity in the Human Brain, NeuroImage, vol. 2, no. 4, pp. 237243, 1995.

[6] L. Gao, W. Cheng, J. Zhang, and J. Wang, EEG classification for motor imagery and resting state in BCI applications using multi-class Adaboost extreme learning machine, Review of Scientific Instruments, vol. 87, no. 8, p. 085110, 2016.

[7] A. Ghaemi, E. Rashedi, A. M. Pourrahimi, M. Kamandar, and F. Rahdari, Automatic channel selection in EEG signals for classification of left or right hand movement in Brain Computer Interfaces using improved binary gravitation search algorithm, Biomedical Signal Processing and Control, vol. 33, pp. 109118, 2017
[8] A. B. M. A. Hossain, M. W. Rahman, M. A. Riheen, "Left and right hand movements EEG signals classification using wavelet transform and probabilistic neural network," International Journal of Electrical and Computer Engineering (IJECE), Vol. 5, No. 1,pp. 92 101, 2015
[9] E. F. Gonzlez-Castaeda, A. A. Torres-Garca, C. A. Reyes-Garca, and L. Villaseor-Pineda, Sonification and textification: proposing methods for classifying unspoken words from EEG signals, Biomedical Signal Processing and Control, vol. 37, pp. 8291, 2017.
[10] ~~http://wavelets.pybytes.com/wavelet/coif2/~~
[11] S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE transactions on pattern analysis and machine intelligence, vol. 11, pp. 674693, 1989.
[12] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, Feature reduction and selection for EMG signal classification, Expert Systems with Applications, vol. 39, no. 8, pp. 74207431, 2012.
[13] B. Schlkopf, A. Smola, and K.-R. Mller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation, vol. 10, no. 5, pp. 12991319, 1998.
[14] T. K. Ho, The random subspace method for constructing decision forests, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832844, 1998.
[15] Lai, M. J. Reinders, and L. Wessels, Random subspace method for multivariate feature selection, Pattern Recognition Letters, vol. 27, no. 10, pp. 10671076, 2006.
[16] M. Skurichina and R. P. W. Duin, Bagging and the random subspace method for redundant feature spaces, Multiple Classifier Systems Lecture Notes in Computer Science, pp. 110, 2001.
[17] http://www.bsp.brain.riken.jp/~qibin/homepage/Datasets.html