International Journal of Computational Intelligence and Informatics, Vol. 4: No. 1, April - June 2014

Swarm Optimized Feature Selection of EEG Signals for Brain-Computer Interface

K Akilandeswari
Department of Computer Science
Government Arts College (Autonomous), Salem-7, India.

G M Nasira
Department of Computer Science
Chikanna Government Arts College, Tirupur, India.

Abstract- **A Brain-Computer Interface (BCI) is a communication system which uses cerebral activity to control external devices or computers. BCI research's goal is to provide communication capability to the people who are totally paralyzed or suffer neurological neuromuscular disorders like amyotrophic lateral sclerosis, brain stem stroke or spinal cord injury. A BCI system records brain signals and applies machine learning algorithms to classify such signals, performing a computer controlled action. This study investigates effects of feature selection for Electroencephalograph (EEG) signals. Feature extractions using Walsh Hadamard Transform (WHT) and feature selection with Principal Component Analysis (PCA) are also studied. Feature selection through Particle Swarm Optimization (PSO) is proposed. Classification of the features is achieved through Bagging and decision tree classifiers.**

Keywords - Brain-Computer Interfaces (BCIs), Electroencephalograph (EEG), Walsh Hadamard Transform (WHT), Bagging

I. INTRODUCTION

Brain–Computer Interfaces (BCIs), also called Brain Machine Interface, are communication systems which help humans to interact using electroencephalographic activity control signals and without involving peripheral nerves and muscles. BCI creates a new non-muscular channel to relay a person's intentions to external devices like speech synthesizers, computers, assistive appliances and neural prostheses. BCI are artificial intelligence systems that recognize patterns set in brain signals in five stages: signal acquisition, preprocessing or signal enhancement, feature extraction, classification, and control interface as depicted in Figure 1. Signal acquisition captures brain signals and perform noise reduction/artifact processing. Preprocessing prepares signals in suitable form for additional processing. Feature extraction identifies brain signals discriminative information that was recorded. Once measured, a signal is mapped to a vector having effective and discriminate features from observed signals. Extraction of this information is challenging. Brain signals are mixed with signals from a finite brain activities overlapping time and space. Also, signals are not stationary and may be distorted by artifacts like electromyography (EMG) or electroocoulography (EOG) [1].

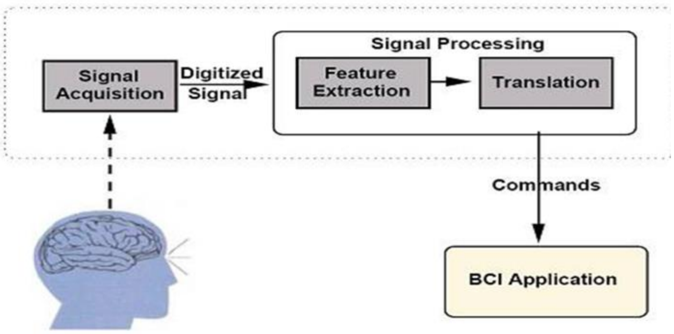


Figure 1: Brain computer interface

A BCI system's main parts are signal acquisition system, feature extractor, feature translator, control interface and device controller. The signal acquisition system involves electrodes which pick up brain electrical activity, amplifier and analog filters. Feature extractor converts brain signals to relevant feature components. At first, EEG

ISSN: 2349 - 6363

raw signals are filtered by digital band pass filter. Then, amplitude samples are squared to get power samples which are averaged for trials. Finally signals are smoothed by averaging over time samples. Feature translator classifies feature components to logical controls. Control interface converts logical controls to semantic controls. Device controller changes semantic controls to physical device commands, different from one device to another based on application. Finally, commands are executed by device [2].

The brain produces electrical and magnetic activity. Hence, sensors detect various types of changes in electrical/magnetic activity at varied times over different brain areas, to study brain activity. BCIs rely on brain activity electrical measures relying on sensors placed on the head to measure activity. Electroencephalography (EEG) is recording electrical activity from scalp with electrodes. It is an established method, used in clinical/research settings for years. EEG equipment is inexpensive, lightweight, and easily applicable. Temporal resolution – the ability to detect changes in a specific time interval is good. Nonetheless, EEG also has disadvantages: Spatial (topographic) resolution and frequency range are limited. EEG is vulnerable to artifacts, which are EEG contaminations caused by electrical activities. Examples are bioelectrical activities due to eye movements/eye blinks and from muscles near recording sites. External electromagnetic sources like power lines also contaminate EEGs. BCI analyzes current brain activity for brain patterns generated from specific brain areas. To ensure consistent recordings from specific head regions, scientists use a standard system to accurately place electrodes, called the International 10–20 System used in clinical EEG recording and EEG and BCI research [3].

EEG is neurophysiologic measurement of the brain's electrical activity using electrodes on the scalp. Resulting traces are called EEG representing an electrical signal (postsynaptic potentials) from many neurons. EEG is a non-invasive procedure used for diagnosis. Instead of electrical currents, voltage differences between various brain sections are observed. The EEG has a set of multi-channel signals. Signals' pattern change reflects large-scale brain activities. Additionally, EEG also reflects head musculature, eye movements activation and also interference from nearby electric devices, and changing electrodes conductivity due to movements of subject or physicochemical reactions at electrode sites[4]. The brain wave of a normal person show different rhythmic activity depending upon his/her level of consciousness. The various rhythms of brain wave are:

Alpha rhythm: Amplitude is variable and mostly below 50 μV in adults. Blocked or attenuated by attention, especially visual/mental effort. Alpha rhythm is temporarily blocked, i.e., its amplitude decreased, by eye opening, other stimuli or mental activities. Reactivity degree varies.

Mu rhythm: Mu rhythm frequency is 10 Hz and amplitude below 50 μV. Mu stands for motor and mu rhythm is related to motor cortex functions and nearby somatosensory cortex. Mu rhythm is blocked by movements/light tactile stimuli. Thoughts about performing movements and readiness to move also block mu rhythm, making it important in BCI research.

Beta rhythms: Any rhythmical activity in 13-30 Hz frequency band is considered beta rhythm whose amplitudes are rarely larger than 30 μV.

BCIs based on rhythmic activity: BCI research considered using hand/foot movement imagination as its basis. Hence, mu rhythm has an important role. Studies stated use of mu rhythm in BCI concluding that "mu rhythm is modulated by expression of self-generated movement and by observation and movement imagination." However, in EEG biofeedback, self-regulation of alpha or beta rhythms was extensively [5] used.

To select appropriate BCI system classifier, it is essential to understand features used, their properties and use. This section describes common BCI features and specifically their properties and how to use them to consider EEG time variations. Various features were attempted to design BCI like EEG signal amplitude values, Band Powers (BP), Power Spectral Density (PSD) values, Auto Regressive (AR) and Adaptive Auto Regressive (AAR) parameters, time-frequency features and inverse model-based features.

Concerning BCI design some critical properties to be considered include:

- Noise and outliers: BCI features are noisy with outliers as EEG signals have poor signal-to-noise ratio;
- High dimensionality: Feature vectors are of high dimensionality in BCI systems. Many features are extracted from many channels and from many time segments prior to concatenation into one feature vector;
- Time information: BCI features have time information as brain activity patterns are related to specific EEG time variations.
- Non-stationary: BCI features are non-stationary as EEG signals rapidly vary over time and sessions;
- Small training sets: Training sets are relatively small, as training process is time consuming and demanding for subjects [6].

These properties are verified for features in BCI research, but not true for BCI in clinical practice. For instance, training sets for a patient would not be small as huge amount of data is acquired during days/months sessions. Usually many domains features combine for better results. For example time domain and band pass filtering precede spatial pattern extraction. This paper investigates feature selection effect for EEG signals using PCA with feature extraction by WHT. In this study, feature selection based on Particle Swarm Optimization (PSO) is proposed and classification is carried out with Bagging with decision tree classifiers. The study is organized as follows: section 2 deals with reviews of BCI approaches in literature. Section 3 discusses methodology used in this study, Section 4 reports results of experiments, and Section 5 concludes the paper.

II. LITERATURE SURVEY

A new, effective S-PCA based feature selection algorithm and wavelet transform features in EEG signal classification was proposed by Nasehi and Pourghassem [7] where a new, effective feature S-PCA and WT features in medical and BCI application was suggested. This decomposed signals to six sub-bands by four mother wavelet. Then five features from each sub-band were extracted as feature vectors. S-PCA selected ten effective features from WT features in this algorithm and used KNN classifier and seven different brain activity signals to evaluate the new method. Results reveal improvement in classification performance compared to current methods.

Feature subset selection and feature ranking for multivariate time series was proposed by Yoon, et al., [8] where the author suggested a family of new, unsupervised methods for feature subset selection from MTS based on PCA, termed clever. Traditional FSS techniques like RFE and FC were applied to MTS data sets [BCI data sets]. Such techniques lose correlation information among features, while the new techniques used PCA properties to retain such information. To evaluate selected features subset effectiveness, it employed classification as target data mining task. Experiments showed that clever outperformed RFE, FC and random selection up to a factor of two regarding classification accuracy, while taking up two orders of magnitude less processing time compared to both RFE and FC.

An efficient words typing P300-BCI system using modified T9 interface and random forest classifier was proposed by Akram, et al., [9], where the author suggested modifying both to increase word typing speed and accuracy. In the paradigm, it modified T9 interface similar to mobile phones keypad used for texting. It then integrated a custom-built dictionary to provide users with word suggestions while typing. Users select one of the suggestions to complete word typing. The new paradigms reduced word typing time and made words typing convenient by typing complete words with few character spellings. The new system ensured an average of 1.83 minutes per word, when typing ten random words, while conventional spelling required 3.35 minutes for same words under similar conditions, decreasing typing time by 45.37%.

Application of covariate shift adaptation techniques in BCI was proposed by Li, et al., [10]. This was applied on a BCI Competition III dataset with results revealing that covariate shift adaptation compared favourably with methods in BCI competition in coping up with non-stationaries. Specifically, bagging with covariate shift increased stability, when applied to competition dataset. An online experiment proved bagged-covariate shift method’s effectiveness. It can thus be summarized that covariate shift adaptation helped realize adaptive BCI systems.

BCI feature selection and classification was proposed by Polat and Cataltepe [11] where the author suggested many features being extracted from raw EEG data before feature selection and classification, for BCI applications using motor imaginary movements. As mRMR feature selection was quick to select relevant and non-redundant feature sets, it was chosen. Using different classifiers, it was seen that feature selection helps classification performance with higher classification accuracy being achieved with less features. BCI Competition 2003 3A data set was used in experiments.

Implementation of automatic feature selection methods for BCI realization was proposed by Majkowski, et al., [12] which suggested that BCI’s main task was translating brain neuron generated signals into commands. For effective BCI operation, efficient EEG signals feature selection was needed. The authors proposed use of correlation and t-statistics for feature selection in this article.

Automatic feature selection for BCI was proposed by Coelho, et al., [13]. An analysis using davies-bould in index and extreme learning machines, it presented a new automatic feature selection framework in BCIs. The proposal manipulated features generated in frequency domain by estimation of power spectral EEG signals density was based on feature optimization with a state-of-the-art artificial immune network, cob-aiNet. To analyze the proposed framework’s performance, two approaches were adopted: direct use of Davies-Bouldin index and metrics associated with operating an ELM as classifier. Results reveal the proposal has potential to improve BCI system performance and provide elements to analyse EEG signals spectral content and ELMs performance in motor imagery paradigms.

Spectral EEG features/tasks selection process considerations in BCI applications was proposed by Dobrea, et al., [14], where the author developed subject specific mental tasks selection process as necessary in EEG-based

BCI applications. While, two previous researches proved - using EEG-extracted AR parameters and 12 varied mental tasks - major gains one gets in tasks classification performance is only through selecting proper tasks. It investigated putative relation between every pair and corresponding individual optimum cognitive tasks set. This set considered three different spectrum relative power parameters.

An improved feature extraction method for self-paced BCI application was proposed by Guangming, et al., [15] where improved feature extraction and selection methods were seen. Feature extraction was by stationary wavelet transform and band pass filtering with SVM classifying extracted feature vectors. This method selected EEG features and classifier parameters through genetic algorithm and tested them in BCI competition 2008 dataset I. Informal results prove the method to be efficient for feature extraction and selection in self-paced BCIs.

Extracting and selecting discriminative features, from high density NIRS-based BCI for numerical cognition was proposed by Ang, et al., [16]. This presented a study with high density 348 channels NIRS-based BCI from 8 healthy subjects when solving mental arithmetic problems with2 difficulty levels and rest condition. Current feature extraction and selection methods on present study were presented for low density 16 channels NIRS-based BCI alone, needing specification on many features to result in a desirable performance. This work presented a discriminative features extraction method from high density single-trial NIRS data with common average reference spatial filtering and single-trial baseline reference, and a method to automatically select a set of discriminative and non-redundant features using MIRSR and SPSEC algorithms. Results showed effectiveness of the proposed feature extraction and selection method in high density NIRS-based BCI to assess numerical cognition.

An algorithm for feature down-selection on AGV based subject was proposed by Dias, et al., [17]. AGV was evaluated and compared with three algorithms RFE, simple GA and RELIEF algorithm. 5 healthy subjects high-dimensional data was reduced by the algorithms under experiment and classified on alternative right hand/foot movement imagery tasks. AGV outperformed other methods simultaneously selecting smallest feature subsets. Dimensionality reduction with high discrimination power was best on AGV's performance.

A combination of amplitude and phase features under uniform framework with EMD in EEG-based BCIs was proposed by He, et al., [18], where the author applied this method to public and recorded datasets. Compared to traditional CSP, average classification accuracy increase is 5.4%, which manifests statistical significances. Also, it investigated chances of online realization of this method showing comparable offline results.

Ang, et al., [19] proposed feature extraction from high density NIRS-based BCI to assess numerical cognition which investigated performance of high density 348 channels NIRS-based BCI on 8 healthy subjects when solving mental arithmetic problems with 2 difficulty levels and rest condition. A new feature extraction method from high density single-trial NIRS data was proposed with common average reference spatial filtering and single-trial baseline reference. The new feature extraction method’s performance was presented using 5×5-fold cross-validations on single-trial NIRS data got using mutual information-based feature selection and SVM classifier. Results demonstrated new method’s effectiveness in high density NIRS-based BCI to assess numerical cognition.

VEP feature extraction and classification for BCI was proposed by He, et al., [20], which revealed that feature extraction in wavelet domain extracted VEP signal features, reduced noise and decreased dimensionality simultaneously and VEP classification algorithms recognized VEP signals producing BCI control signals correctly to increase BCI’s information transfer rate.

III. METHODOLOGY

A. Dataset

The dataset that evaluated the new method is Data Set I for BCI Competition III. In BCI experiments, a subject performs imagined movements of left small finger or tongue. Electrical brain activity’s time series is picked up in the trials. Recordings had a sampling rate of 1000Hz. Recorded potentials were stored after amplification, as microvolt values. Each trial had either an imagined tongue or imagined finger movement recorded for 3 seconds. To prevent data reflecting visually evoked potentials, recording intervals started 0.5 seconds after conclusion of visual cue. Brain activity in 278 trials was training data and 100 trials brain activity was test data [21].

The flowchart of the methodology is shown in figure 2.

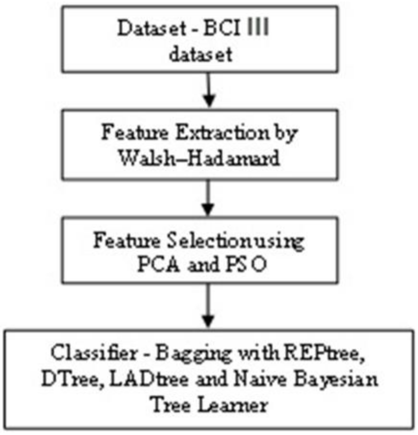


Figure 2: Flowchart of the Methodology

B. Feature Extraction by Walsh–Hadamard

The Walsh transform/Walsh–Hadamard transform is a non-sinusoidal, orthogonal transformation technique decomposing signals to basic functions which are Walsh functions rectangular or square waves with values of +1 or –1. Walsh–Hadamard transform returns sequency values. Sequency is a generalized frequency notion defined as one half of average zero-crossings per unit time interval. Each Walsh function has a unique sequency value. Walsh-Hadamard transform is used in many applications like image processing, speech processing, filtering and power spectrum analysis. It reduces bandwidth storage and spread-spectrum analysis. Like Fast Fourier Transform (FFT), Walsh–Hadamard transform has a fast version, Fast Walsh-Hadamard Transform (FWHT). Compared to FFT, FWHT needs less storage space and is quicker in calculations as it uses only additions and subtractions, while FFT needs complex values. FWHT represents signals with sharp discontinuities accurately using lesser coefficients than FFT. FWHT is a divide and conquer algorithm recursively breaking down WHT of size N to two smaller WHTs of size N / 2. This implementation follows recursive definition of 2N X 2N Hadamard matrix HN [22].

$$H_N = \frac{1}{\sqrt{2}}\begin{pmatrix} H_{N-1} & H_{N-1} \\ H_{N-1} & -H_{N-1} \end{pmatrix} \quad (1)$$

For discrete series of length N, a set of Walsh functions is N by N Hadamard matrix. Whereas Fourier analysis efficiently describes signals of highly localized frequency components Walsh functions suit broadband signal components and so were used in applications like speech coding to describe broadband components efficiently while Fourier coefficients [23] described narrow band components.

C. Feature Selection

Feature extraction is applied to raw data and a classifier is trained using all extracted features. The problem is that it uses all extracted features for classification which results in the use of irrelevant information being fed to the classifier. This disallows the classifier from proper generalization, thus feature selection is used to extract a subset of features which are used for classification. In this paper, features extracted from the EEG are selected using Principal Component Analysis (PCA) and a proposed swarm based feature selection. The proposed feature selection is based on PSO.

D. Principal Component Analysis

PCA is a valuable result from applied linear algebra. PCA is used in all analysis from neuroscience to computer graphics as it is a simple, non-parametric method to extract relevant information from confusing data sets. PCA is an orthogonal linear transformation converting data to a new coordinate system like greatest variance by any data projection which lies on first coordinate, second greatest variance on second coordinate and so on. If X^T with zero empirical mean, where each of n rows represents varied experiment repetition and each of m columns ensures specific datum. Singular value decomposition of X is X = WΣVT, where m × m matrix W is matrix of eigenvectors of covariance matrix XXT, matrix Σ is a m × n rectangular diagonal matrix with nonnegative real numbers on diagonal, and n × n matrix V is matrix of eigenvectors of XTX.

E. Particle Swarm Optimization (PSO)

PSO is used to find the optimal feature subset of the Eigen features extracted using PCA. The proposed method finds an optimal feature subset with least number of features with high classification accuracy. Every particle flies in search space in PSO with a velocity adjusted by own flying memory and companion's flying experience. Each particle has objective function value decided by fitness functions:

$$v_{id}{}^{t} = W \times v_{id}{}^{t-1} + c_1 \times r_1(p_{id}{}^{t} - x_{id}{}^{t}) + c_2 \times r_2(p_{gd}{}^{t} - x_{id}{}^{t}), \quad (2)$$

Where i represents i_{th} particle and d is solution space dimension, c1 denotes cognition learning factor, and c2 indicates social learning factor, r1 and r2 are random numbers distributed uniformly in (0,1), pid ^{t}and pgd^{t}stands for position with best fitness found till then for ith particle and best position in neighbourhood, vid^{t}and vid^{t-1}are velocities at time t and time $t-1$, respectively, and xid ^{t} is position of ith particle at time t.

Each particle moves to a new potential solution based on the equation [24]:

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t}, d = 1,2 \ldots . D \quad (3)$$

$$x_{id} = \begin{cases} 1, & rand() < s(V_{id}) \\ 0 \end{cases} \quad (4)$$

$$s(v) = \frac{1}{1+e^{-v}} \quad (5)$$

The particles of the PSO represent the features subset where if a bit is assigned value 1 then the feature is selected and if assigned 0, it is not. The particles fly through the feature space and during the iteration of the algorithm, converge to optimal position. The fitness function of the particles is evaluated based on the classification accuracy and the number of features. The fitness is given as:

$$fitness = \alpha * accuracy + \beta * number\ of\ features$$

Where α, β are weights assigned to accuracy and number of features and α + β =1. In this study, as the importance is on classification accuracy, α is assigned higher weightage of 0.8 and β=0.2. The binary mapping is used for PSO mapping and is shown in the table 1.

TABLE I. PSO MAPPING

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
T1	0	1	0	0	1	0	0	1	0	0
T2	0	0	1	1	0	1	0	1	1	1
T3	0	0	0	1	0	1	0	0	1	1
T4	1	1	1	1	1	1	1	1	1	1

Now the features become the initial particles, where 0 represents feature not selected and 1 representing feature selected. So, PSO randomly takes features and compute error and becomes the pbest. The lowest error becomes the gbest. Velocity is computed and PSO work goes on.

F. Classifiers

1) Bagging

Bagging technique introduced by Breiman [25] uses bootstrap samples from original dataset to build replicate classifiers. Final classification is through a combination that weights individual classifiers outputs. If Sn = (xi, yi), i=1,...,nis training set, then n number of instances are replaced to create a bootstrapped sample. It is of similar size as original training data, with duplicates of some instances, leaving out some original data instances. Usually, bootstrap sampling is executed between 25 to 50 times. Traditional machine learning technique trains each training samples. Bagged estimate is found by averaging resulting estimator as

$$f(x) = \sum_{i=1}^{m} \alpha_i h_i(x) \quad (7)$$

Where f(x) is resulting ensemble classifier, hi- base classifier trained on bootstrap sample I and αi- averaging constant.

G. Alternate Decision Trees (ADTrees)

Alternate Decision Trees (ADTrees) [26] are similar to Classification And Regression Trees (CART), differing only in construction / interpretation. ADTrees advantage is that it avoids large and complicated decision trees while maintaining some structured knowledge. ADTrees provide an extra structure layer for a set of weak learners from boosting algorithm. Alternating decision and prediction nodes layers are built in ADTrees. Attributes are identified by prediction nodes and two leaves branching out from it represent attribute domain partition. During unknown instance prediction only one path from a prediction node is followed. Many arbitrary prediction node children and a score are stored in decision nodes. During evaluation, an instance follows a parallel path through all children reaching node. ADTrees start with decision nodes, and a results set obtained by path at each prediction node reached. Scores of decision nodes in the path are summed up and assigned to an instance.

H. REPTree

Decision/regression tree are rapidly built in REPTree. It is a quick decision tree learner using information gain as splitting criterion and reduced error pruning method prunes trees. Numeric attributes values are sorted once. C4.5's method of fractional instances deals with missing values.

I. Naive Bayesian Tree Learner (NBTREE)

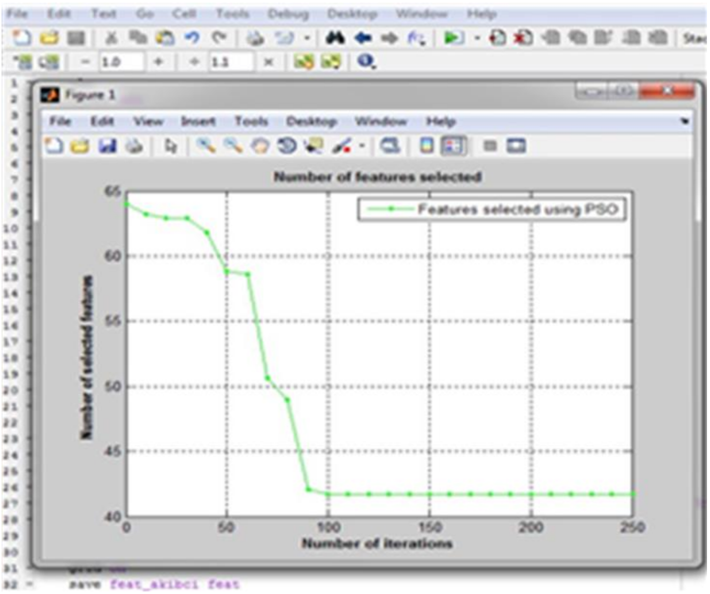
Naive Bayesian tree learner, NBTree [27], is a combination of Naive Bayesian classification and decision tree learning. Local naive Bayes is put on each decision tree leaf in NBTree process. Unknown instance is classified using local naive Bayes on leaf with leaf related data. Hence, an instance sorted to leaf, class label is assigned by Naïve Bayes.

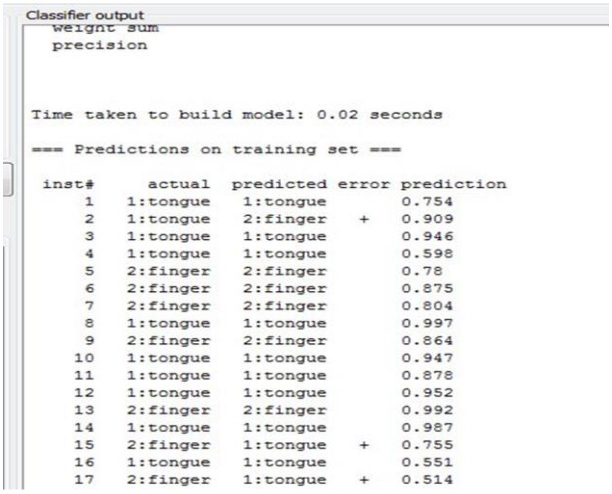
J. LADTree

Using logiboost strategy, a multiclass alternating AD tree known as LADTree is generated. It has the ability to have more than 2 class inputs. Using the Logiboost Strategy additive logistic regression is performed.

IV. RESULTS AND DISCUSSION

Features were extracted using WHT from the EEG and the Maximum and Average energy are used as features to classify the instance. Features were selected using PCA and the proposed PSO. The efficacy of the feature selection is evaluated. Experiments were conducted with various classifiers and the improvement in the classifier due feature selection is assessed based on parameters like classification accuracy, recall and precision. Initially 64 features were extracted from EEG. On applying PCA technique, 45 features were considered while remaining features re removed. The proposed PSO selected 38 features for classification. Training set was made up of 60% of the data and the remaining 40% was used as test set. Figure 3 and 4 is the features selected using PSO and classifier output obtained respectively.





```
Classifier output
  weight sum
  precision


Time taken to build model: 0.02 seconds

=== Predictions on training set ===

 inst#     actual   predicted error prediction
     1   1:tongue   1:tongue         0.754
     2   1:tongue   2:finger    +    0.909
     3   1:tongue   1:tongue         0.946
     4   1:tongue   1:tongue         0.598
     5   2:finger   2:finger         0.78
     6   2:finger   2:finger         0.875
     7   2:finger   2:finger         0.804
     8   1:tongue   1:tongue         0.997
     9   2:finger   2:finger         0.864
    10   1:tongue   1:tongue         0.947
    11   1:tongue   1:tongue         0.878
    12   1:tongue   1:tongue         0.952
    13   2:finger   2:finger         0.992
    14   1:tongue   1:tongue         0.987
    15   2:finger   1:tongue    +    0.755
    16   1:tongue   1:tongue         0.551
    17   2:finger   1:tongue    +    0.514
```

Figure 4: Classifier Output Obtained

The WEKA tool is used to assess the performances of the classification algorithms with their default parameter setting. Bagging with REPtree, ADTree, LADTree and Naïve Bayes tree are all tested. All the classification performances are based on the 10-fold cross validations for a more accurate evaluation. The experimental results and the classification accuracy are obtained. The results obtained are shown in Figure 5-8.

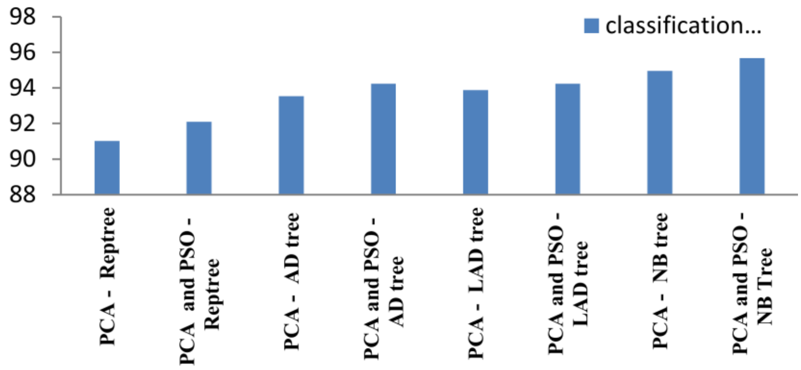


Figure 5: Classification Accuracy

Fig. 5 show that the proposed WHT with PCA and PSO - Bagging with Naïve Bayes classifier produces a high classification accuracy of 95.68%.

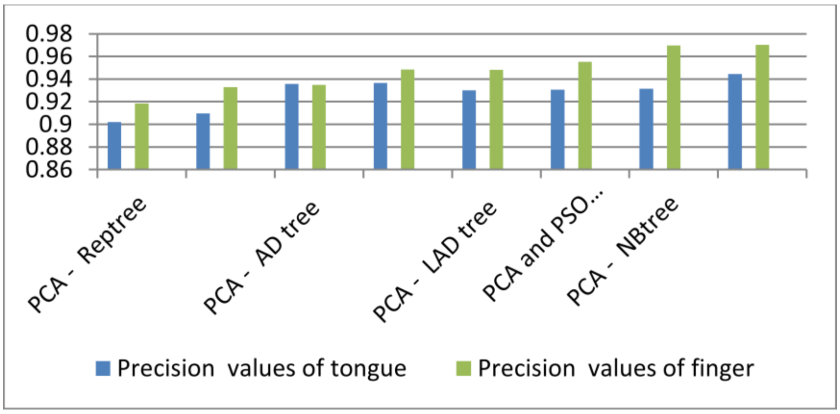


Figure 6: Precision value

From Fig. 6, it is shown that the proposed WHT with PCA and PSO - Bagging with Naïve Bayes classifier provide precision of 94.4 % for tongue and 97 % for finger.

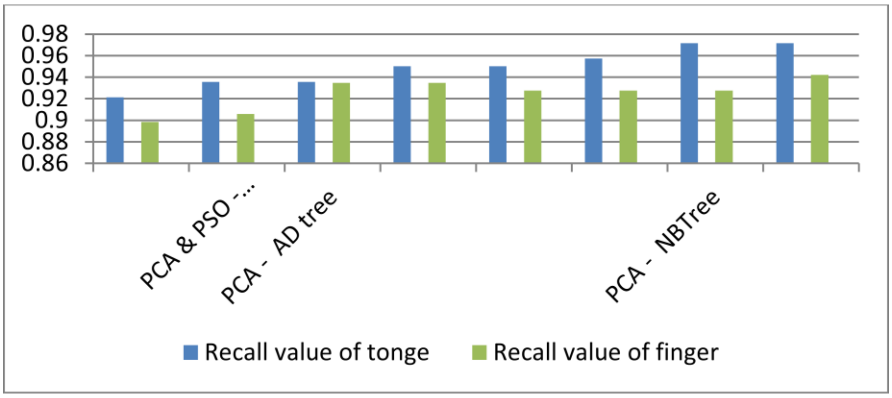


Figure 7: Recall Value

From Fig.7 it is shown that the proposed WHT with PCA and PSO - Bagging with Naïve Bayes classifier provide recall of 97.14% for tongue and 94.2 % for finger.

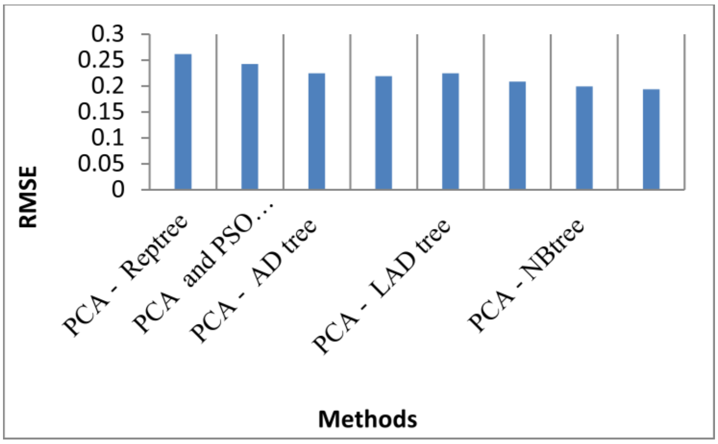


Figure 8: Root Mean Square Error

Fig. 8 show that the proposed WHT with PCA and PSO - Bagging with Naïve Bayes classifier produces a low RMSE value of 1.93 %.

V. Conclusion

BCI systems discriminate between varied brain signal patterns accurately so that users perform various mental tasks. This study investigates using WHT to convert EEG signals from time domain to frequency domain for feature extraction for classification using Bagging. A feature selection based on PSO is also proposed. Experiments undertaken with Data Set I for BCI Competition III were carried out through 10 fold cross validation and accuracy achieved was comparable to results from other research in literature. The results show that WHT with proposed PSO feature selection and Bagging with Naïve Bayes classifier achieves high classification accuracy of 95.68%.This method has very fast feature extraction/classification, but additional work is needed to improve classification accuracy.

References

[1] L. F Nicolas-Alonso, & J. Gomez-Gil, "Brain computer interfaces, a review," Sensors, 12(2),pp. 1211-1279, 2012.

[2] H.H. Alwasiti, I. Aris, & A. Jantan, "Brain Computer Interface Design and Applications: Challenges and Future," World Applied Sciences Journal, 11(7), pp. 819-825, 2010.

[3] Graimann, B, "Brain-computer interfaces," Springer, 2010.

[4] V. Bartošová, O. Vyšata, & A. Procházka, "Graphical user interface for EEG signal segmentation," In Proc. of 15th Annual Conf. Technical Computing, Prague , 22, pp. 1-6, 2007.

[5] http://www.thedirectdata.com/seminar/braininterfacecomputer.doc.

[6] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, & B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," Journal of neural engineering, 4, 2007.

[7] S. Nasehi, & H. Pourghassem, "A novel effective feature selection algorithm based on S-PCA and wavelet transform features in EEG signal classification," In Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference, pp. 114-117, 2011.

[8] H. Yoon, K. Yang, & C. Shahabi, "Feature subset selection and feature ranking for multivariate time series," Knowledge and Data Engineering, IEEE Transactions, 17(9), pp.1186-1198, 2005.

[9] F. Akram, H.S. Han, H.J.Jeon, K. Park, S.H. Park, J. Cho, & T.S Kim, "An efficient words typing P300-BCI system using a modified T9 interface and random forest classifier," In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 2251-2254, 2013.

[10] Y. Li, H. Kambara, Y. Koike, & M. Sugiyama, "Application of covariate shift adaptation techniques in brain–computer interfaces," Biomedical Engineering, IEEE Transactions, 57(6), 1318-1324, 2010.

[11] D. Polat, & Z. Cataltepe, "Feature selection and classification on brain computer interface (BCI) data," In Signal Processing and Communications Applications Conference (SIU), pp. 1-4, 2012.

[12] A. Majkowski, M. Kolodziej, & R.J. Rak, "Implementation of automatic feature selection methods for BCI realization," In Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International pp. 1286-1289, 2012.

[13] G.P. Coelho,C.C Barbante, L. Boccato, R.R. Attux, J.R. Oliveira, & F.J. Von Zuben, "Automatic feature selection for BCI: An analysis using the davies-bouldin index and extreme learning machines," In Neural Networks (IJCNN), The 2012 International Joint Conference, pp. 1-8, 2012.

[14] M. Dobrea, D.M. Dobrea, & D.Alexa, "Spectral EEG features and tasks selection process: Some considerations toward BCI applications," InMultimedia Signal Processing (MMSP), 2010 IEEE International Workshop, pp. 150-155, 2010.

[15] C. Guangming, Z. Jiacai, & Y. Li, "An improved feature extraction method for self-paced brain-computer interface application," In Complex Medical Engineering, 2009. CME. ICME International Conference, pp. 1-6, 2009.

[16] K.K Ang, J. Yu, & C. Guan, "Extracting and selecting discriminative features from high density NIRS-based BCI for numerical cognition," In Neural Networks (IJCNN), The 2012 International Joint Conference , pp. 1-6, 2012.

[17] N.S. Dias, L.R. Jacinto, P.M. Mendes, & J.H. Correia, "Feature down-selection in Brain-Computer Interfaces," In Neural Engineering, 2009. NER'09. 4th International IEEE/EMBS Conference, pp. 323-326,2009.

[18] W. He, P. Wei, Y. Zhou, & L. Wang, "Combination of amplitude and phase features under a uniform framework with EMD in EEG-based Brain-Computer Interface," In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pp. 1687-1690, 2012.

[19] K.K Ang, J Yu, & C. Guan, "Extracting effective features from high density nirs-based BCI for assessing numerical cognition," In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference, pp. 2233-2236), 2012.

[20] Q. He, B. Wu, H. Wang, & L. Zhu, "Vep feature extraction and classification for brain-computer interface," In Signal Processing, 2006 8th International Conference, 4,2006.

[21] Thomas Lal, ThiloHinterberger, Guido Widman, Michael Schröder, Jeremy Hill, Wolfgang Rosenstiel, Christian Elger, Bernhard Schölkopf, NielsBirbaumer "Methods Towards Invasive Human Brain Computer Interfaces," Advances in Neural Information Processing Systems.

[22] D.H. Kekre, V. Kulkarni, S. Venkatraman, A. Priya, &S. Narashiman, "Speaker Identification using Row Mean of DCT and Walsh Hadamard Transform," International Journal on Computer Science and Engineering, 3(1),2011.

[23] J.J Wells, "Real-time spectral modelling of audio for creative sound transformation," (Doctoral dissertation, University of York) 2006.

[24] D. Jin, D. He, D. Liu, & C.Baquero, "Genetic algorithm with local search for community mining in complex networks," In Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on , 1, pp. 105-112, 2010

[25] L. Breiman, Bagging predictors. Machine Learning, 24(2), pp.123–140, 1996.

[26] Y. Freund, L. Mason, "The alternating decision tree learning algorithm," Proceeding of the Sixteenth International Conference on Machine Learning, Bled, Slovenia, pp.124-133, 1999.

[27] R. Kohavi, "Scaling up the cccuracy of naive-Bayes classifiers: a decision-tree hybrid," Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), AAAI Press, pp. 202-207, 1996.