ISSN 2229-5518

A Hybrid Meta heuristic feature selection algorithm for Brain Computer Interface

K.Akilandeswari , G.M.Nasira

Abstract-A Brain-Computer Interface (BCI), also called Brain Machine Interface (BMI), is a communication system allowing users to interact with electronic devices through control signals from Electroencephalographic (EEG) activity sans engaging peripheral nerves and muscles. Preliminary BCI research motivation was to develop assistive devices for those with locked-in disabilities. Currently, researchers are exploring BCI as a new anthropomorphic interaction channel for applications like robotics, virtual reality and games. This paper Investigate the effect of feature selection in BCI. Since feature selection is NP Hard, a novel feature selection technique using Particle Swarm Optimization (PSO) is proposed. The proposed feature selection technique shows improvement in the classification accuracy compared to Principal Component Analysis (PCA). Since meta heuristic algorithms are known to suffer from local minima problem, a hybrid PSO using the principles of hill climbing algorithm is proposed which improved the classification accuracies further.

Index Terms-Brain Computer Interface (BCI), Feature Selection, Principal Component Analysis (PCA), Particle Swarm Optimization (PSO).

1 INTRODUCTION

Brain-Computer Interface (BCI) is a hardware and software communications system allowing cerebral activity to control computers or external devices. BCI research's goal is ensuring communications capabilities to the severely disabled, totally paralyzed or 'locked in' by neurological neuromuscular disorders, like amyotrophic lateral sclerosis, brain stem stroke, or spinal cord injury [1]. BCI activates electronic or mechanical devices through brain activity. BCIs allows direct brain communication in totally paralyzed patients and aims to restore movement in paralyzed limbs through transmitting brain signals to muscles or external prosthetic devices [2]. There are many BCI phases like [3] (Figure 1):

Signal Acquisition
Signal Pre-Processing
Signal Classification
Computer Interaction.

Electroencephalographic (EEG) records the brain's electrical activity. EEG is a variation of electrical fields in the cortex or on the scalp surface caused by the brain's physiological activities

• K.Akilandeswari is working as Associate professor in Department of Computer Science, Government Arts College (Autonomous), Salem-7, India. E-mail : akilamanivelan@gmail.com

• G.M.Nasira is working as Assistant Professor of Computer Applications in Department of Computer Science , Chikanna Government Arts College, Tirupur, India. E-mail : nasiragm99@yahoo.com

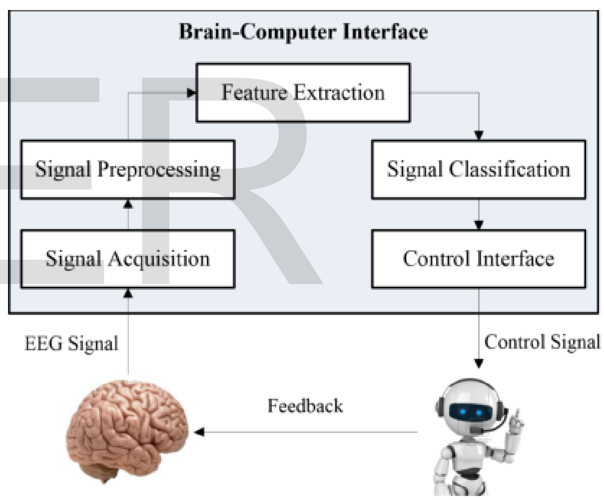

Figure 1 Schematic of a BCI system

EEG is a widely adopted method to assess brain activity. Detecting wave changes is critical to understand brain function. Spontaneous EEG signals in clinical applications are divided into many rhythms according to frequency [4].

EEG is a graphic representation of voltage differences between two different cerebral locations plotted over time. A scalp EEG signal generated by cerebral neurons is modified by tissues electrical conductive properties between electrical source and recording electrode on a scalp, the electrode's conductive properties and the cortical generator's orientation to the recording electrode. EEG is obtained as the process of current flow through tissues between electrical generator and recording electrode is called volume conduction. EEG ensures a two-dimensional projection of three-dimensional reality, meaning that theoretically it is impossible to

ISSN 2229-5518
determine the EEG generator location based on scalp-recorded EEG information alone [5].

Feature Selection (FS) is a machine learning global optimization problem which reduces features, removes irrelevant, noisy, and redundant data resulting in acceptable recognition accuracy. It is an important step affecting a pattern recognition system performance. Feature selection can be a pre-processing tool of importance before solving classification issues. Feature selection aims to reduce maximum irrelevant features while maintaining acceptable classification accuracy. Feature selection is important in pattern classification, multimedia information retrieval, data analysis, remote sensing, biometrics, machine learning, computer vision, medical data processing, and data mining applications.

This work presents a PSO based feature selection for EEG signals, feature extraction using WHT and feature selection using PCA. Classification is through Bagging with Naïve Bayes classifiers. The remainder of the paper is organized as follows: Section 2 reviews some related works in literature, section 3 explains techniques used in the investigation, section 4 presents results and discussion and section 5 concludes the paper.

2 LITERATURE REVIEW

A new PSO based feature selection method for EEG-based Motor-Imagery (MI) BCI systems proposed by Zhiping et al., [6] has two steps: (1) optimization algorithm, i.e. PSO selects EEG features and classifier parameters; and (2) voting mechanism to remove redundant features produced by the optimization algorithm. It compared the new method with a GA method. Experiment on single-trial MI EEG classification revealed the proposed method's effectiveness.

Three feature subsets obtained by tracks extraction method that are wavelet and fractional Fourier transform were compared by Guerrero-Mosquera et al., [7]. It compared each subset's classification tasks performance using SVM and chooses a features combination through forward–backward procedure based feature selection methods and mutual information. Results confirm that fractional Fourier transform coefficients ensured very good performance and that combining features improved classifier performance.

An optimal feature extracting method from the EEG signals basic power density was proposed by Chum et al., [8]. Simulation used a BCI competition III, IV dataset and data experimented on, in a lab. To improve the proposed feature extraction method, it applied extracted feature to a SVM.

Feature reduction and classifier structures were investigated by LaRocco et al., [9]. This paper presents one feature corresponding to maximum average distance between events and non-events, on unbalanced data produced a phi correlation of 0.94 on a mock data with a SNR of 0.3, compared to a phi coefficient of 0.00 for PCA. Simulation proved potential compared to other feature selection or reduction methods.

A master-worker implementation of two different parallel evolutionary models, parallel cost functions computation for population individuals and parallel execution of evolutionary multi-objective procedures on subpopulations proposed by Kimovski et al., [10] showed benefits of parallel processing not only to decrease running time, but also to improve solution quality.

A new feature selection based on Statistical-Principal Component Analysis (S-PCA) and Wavelet Transform (WT) features in medical and BCI applications was proposed by Nasehi and Pourghassem [11]. The S-PCA algorithm selects 10 effective features from WT features. It uses a kNN classifier and 7 different brain activity signals to evaluate the new method. Results reveal classification performance improvement compared to current methods.

A statistically-motivated electrode and feature selection procedure, based on Cohen's effect size f2was proposed by Jenke et al., [12] compared inter and intra-individual selection on a self-recorded database. Classification was evaluated through Quadratic Discriminant Analysis (QDA). While accuracies of up to 57.5% (5 classes) were reached by applying intra-individual selection, inter-individual analysis successfully located features that performed with lower variance in recognition rates on subjects than combinations of electrodes and features as seen in literature.

3 METHODOLOGY

In this work the brain signals are converted to frequency domain using Walsh Hadamard Transform and features extracted from it are selected using PCA. A new hybrid PSO based feature selection is proposed for feature selection. The selected features are classified using Bagging with Naïve Bayes classifiers. The subsequent sections detail each procedure in detail.

3.1 Dataset

BCI Competitions are organized to foster development of improved BCI technology by giving an unbiased validation of various data analysis techniques. A data set is a brain signals record from BCI experiments in leading BCI technology labs split into 2s: one part is labeled data ('training set') and the other unlabeled data ('test set'). Researchers globally could tune their methods to training data and submit their translation algorithms output for test data. The truth about test data was kept secret till a deadline was completed and was used to evaluate submissions. This guarantees that performance assessment is not biased by overfitting selection methods and choice of data parameters.

The dataset used to evaluate the new method is Data Set I from BCI Competition III. A subject performs imagined movements of left small finger or tongue in BCI

ISSN 2229-5518

experiments. Recordings were made at rate of 1000Hz. The recorded potentials were kept as microvolt values after amplification. All trial had 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 visual cue end.

3.2 Feature Extraction

Walsh–Hadamard Transform (WHT)

Walsh–Hadamard Transform (WHT) is a non-sinusoidal, orthogonal transformation technique which decomposes signals to basic functions which are Walsh functions rectangular or square waves with values of +1 or –1. WHTs return sequence values. Sequency is a generalized frequency notion and is one half of average zero-crossings per unit time interval. Each Walsh function has a sequency value. WHT is used in applications like speech processing, image processing, filtering, and power spectrum analysis. It reduces bandwidth storage and spread-spectrum analysis.

WHT like Fast Fourier Transform (FFT), has a fast version, Fast WHT (FWHT). Compared to FFT, FWHT requires reduced storage space and calculates quicker as only additions and subtractions are used while FFT needs complex values. FWHT represents signals accurately with sharp discontinuities using lesser coefficients than FFT. FWHT is a divide and conquer algorithm breaking down WHT of size N to two smaller WHTs of size N / 2. Implementation follows recursive definition of 2N X 2NHadamard matrix HN [22].

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

For discrete series of length N, a Walsh functions set is N by N Hadamard matrix while Fourier analysis describes signals of highly localized frequency components efficiently.

3.3 Feature Selection

Generally, Feature extraction is applied to raw data, and a classifier is trained to use extracted features. The problem is that it uses extracted features for classification results and also uses irrelevant information being fed to a classifier. So feature selection is to be refined to extract a features subset for classification. This work presents EEG extracted features being selected using PCA and a new hybrid PSO based feature selection.

Principal Component Analysis (PCA)

PCA is a valuable result from applied linear algebra used in analysis from neuroscience to computer graphics as it is simple, and non-parametric to extract relevant information from confusing data sets.PCA is an orthogonal linear transformation that converts data to a new coordinate system like greatest variance by a data projection on first coordinate, second greatest variance on second coordinate and so on. XT with zero empirical mean, where each n row represents varied experiment repetition and each 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.

3.4 Particle Swarm Optimization (PSO)

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

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

Where i represents ith 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} are 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. A particle moves to a new potential solution based on equation [24]:

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t}, d = 1,2 \dots . D$$

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

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

The sigmoidal function is used to trash the value between 0 and 1 as the features are represented by binary values with 0 representing the feature not selected and 1 representing the feature as selected. PSO particles represent a features subset where if a bit is assigned a value 1 then feature is selected and if assigned 0, it is not. Particles fly through feature space and during the algorithm's iteration converge to an optimal position. The particle's fitness function is evaluated based on classification accuracy and features.

Fitness is given as:

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

where α, β are weights assigned to accuracy and features and α + β =1. In this investigation, due to the importance of classification accuracy, α is assigned higher weightage of 0.8 and β=0.2.

3.5 Hybrid PSO

Hybrid PSO algorithm starts with an initial K particles swarm. Every particle vector corresponds to the underlying problem's candidate solution. Then, all particles repeatedly move till a maximal number of iterations are passed. During every iteration, a particle individual best and swarm's best positions are determined. A particle adjusts its position based on individual experience (pbest) and swarm's intelligence (gbest) as seen in equations. To expedite convergence speed, all particles are updated using a hill-climbing heuristic before entering next iteration. When the algorithm is terminated, an incumbent gbest and corresponding fitness value are output and considered as optimal task assignment and minimum cost.

$$v_{ij} \leftarrow v_{ij} + c_1 rand_1 (pbest_{ij} - particle_{ij}) + c_2 rand_2 (\mathrm{gbest}_j - \mathrm{particle}_{ij})$$

$$\mathrm{particle}_{ij} \leftarrow particle_{ij} + v_{ij}$$

The flow of the proposed hybrid algorithm is given below.

1. Initialize.
   1.1 Generate K particles at random.
   1.2 Generate velocities vij, $1 \le i \le K$ and $1 \le j \le r$, where vij is randomly drawn from [0.0, 1.0].
2. Repeat until a given maximal number of iterations is achieved.
   2.1 Evaluate the fitness of each particle.
   2.2 Determine the best vector pbest visited so far by each particle.
   2.3 Determine the best vector gbest visited so far by the whole swarm.
   2.4 Update velocities vij using (1) restricted by a maximum threshold vmax.
   2.5 Update particles, vectors using (2).
   2.6 Improve the solution quality of each particle using the embedded hill-climbing heuristic.

3.6 Bagging

Bagging technique introduced by Breiman [13] uses bootstrap samples from original dataset to build duplicate classifiers. Final classification is by a combination that weights individual classifiers outputs. If Sn = (xi, yi), i=1,...,n is training set, then n instances are replaced to create a bootstrapped sample. It is of usually executed between 25 to 50 times. Traditional machine learning techniques train every training sample.Bagged estimate is found by averaging resulting estimator as

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

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

Naïve Bayes classifiers are Bayes theorem [14] based statistical classifiers using a probabilistic approach to predict a data's class, by matching data to a class with highest posterior probability.

4. RESULTS AND DISCUSSION

The experiments are conducted using Data Set I from BCI Competition III. The features are extracted using WHT from the EEG signals. The proposed PSO and hybrid PSO feature selection methods are compared with PCA. The selected features are classified using Naïve Bayes. The experimental results for classification accuracy, precision and recall are tabulated in table2.

Technique	Accu-racy %	Precision	Recall	RMSE
WHT with PCA Bagging with NB tree	94.96	0.9506	0.9495	0.1989
WHT with PCA and PSO -Bagging with NB	95.68	0.9573	0.9567	0.1934
WHT with PCA and Hybrid PSO - Bagging with NB	95.83	0.9586	0.9578	0.1902

From figure 2 it can be seen that classification accuracy improves by 0.91% compared to technique without feature selection. Though the improvement in the classification accuracy has not improve significantly, the processing time comes down due to lower number of features which is critical when implementing in an embedded system. Figure 4 shows the precision and recall with marked improvements.

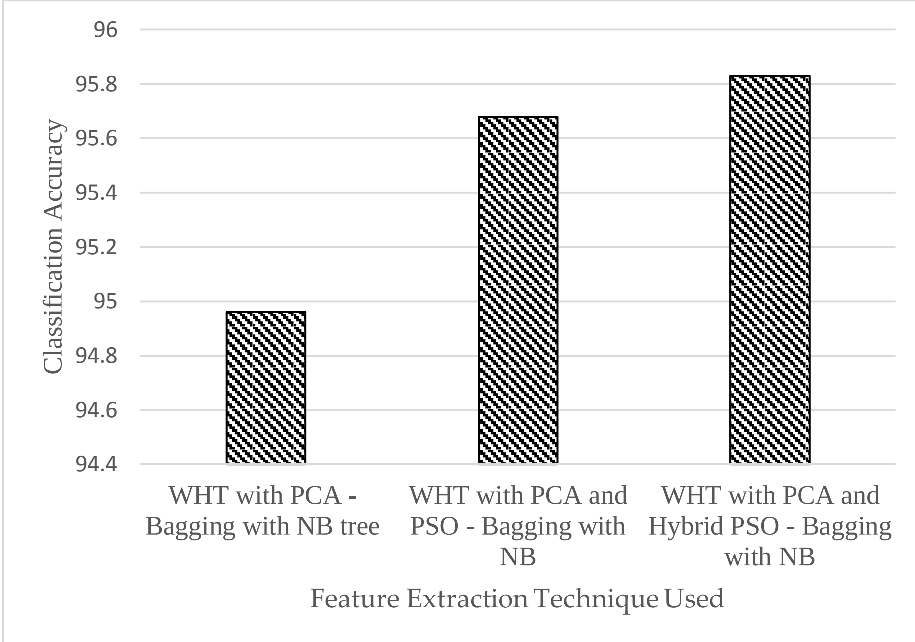

Figure 3 : Classification Accuracy obtained by various techniques.

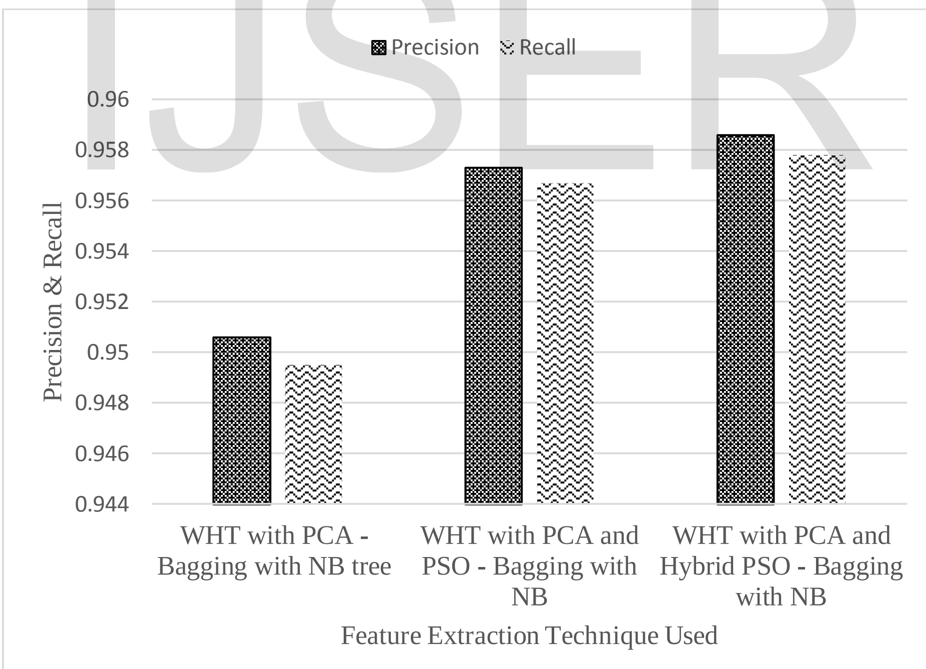

Figure 4 : Precision and recall obtained by various techniques.

Figure 5 shows the Root Mean Squared Error (RMSE). The proposed feature selection technique decreases the RMSE by 4.47% which is statistically significant.

ISSN 2229-5518

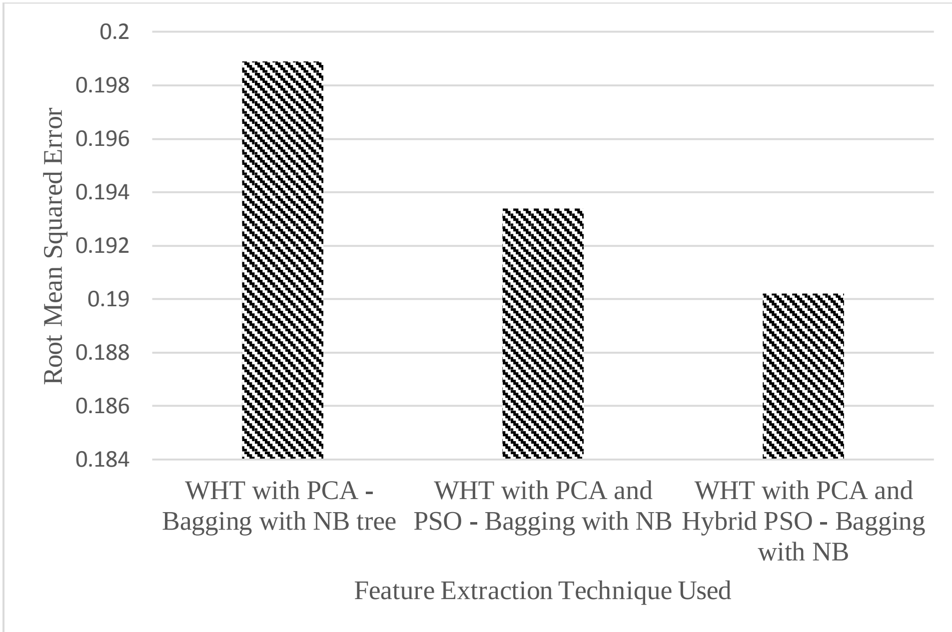

Figure 5 : Root mean squared error.

The decreased RMSE indicates that classification capability of the classifier is improved which is crucial for newer test data where noise distribution could be different.

5. CONCLUSION

This work proposed a novel technique for feature selection in Brain Computer Interface using PSO and hybrid PSO. The proposed hybrid PSO used a sequence of two optimization techniques where in the initial phase of the algorithm uses PSO for feature selection. The selected features are further optimized using Hill Climbing algorithm. The hill climbing algorithm initiates a local search to find a better solution by incrementally improving every available solution. The proposed hybrid PSO solution converged faster compared to traditional PSO and improved the feature set selection. Experiments were undertaken on a Data Set I from BCI Competition III. The proposed technique significantly improved the performance in both the feature selection phase and the classifier phase.

ACKNOWLEDGMENT

We wish to thank Tamilnadu State Council for Higher Education(TANSCHE) for the grant released under Minor Research Project scheme for this work.

REFERENCES

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

[2] Birbaumer, N. (2006). Breaking the silence: brain–computer interfaces (BCI) for communication and motor control. Psychophysiology, 43(6), 517-532.

[3] Rao, T. K., Lakshmi, M. R., & Prasad, T. V. (2012). An exploration on brain computer interface and its recent trends. arXiv preprint arXiv:1211.2737.

[4] Ahmed, S. A., Rani, D. E., &Sattar, S. A. (2012). Alpha Activity in EEG and Intelligence. Mental, 10, 100.

[5] Olejniczak, P. (2006). Neurophysiologic basis of EEG. Journal of clinical neurophysiology, 23(3), 186-189.

[6] Zhiping, H., Guangming, C., Cheng, C., He, X., &Jiacai, Z. (2010, November). A new EEG feature selection method for self-paced brain-computer interface. InIntelligent Systems Design and Applications (ISDA), 2010 10th International Conference on (pp. 845-849). IEEE.

[7] Guerrero-Mosquera, C., Verleysen, M., & Vazquez, A. N. (2010, August). EEG feature selection using mutual information and support vector machine: A comparative analysis. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE (pp. 4946-4949). IEEE.

[8] Chum, P., Park, S. M., Ko, K. E., &Sim, K. B. (2012, October). 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 (pp. 1561-1566). IEEE.

[9] LaRocco, J., Innes, C. R., Bones, P. J., Weddell, S., & Jones, R. D. (2014, August). Optimal EEG feature selection from average distance between events and non-events. In Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE (pp. 2641-2644). IEEE.

[10] Kimovski, D., Ortega, J., Ortiz, A., &Banos, R. (2014, September). Feature selection in high-dimensional EEG data by parallel multi-objective optimization. In Cluster Computing (CLUSTER), 2014 IEEE International Conference on (pp. 314-322). IEEE.

[11] Nasehi, S., &Pourghassem, H. (2011, May). A novel effective feature

ISSN 2229-5518
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 on (pp. 114-117). IEEE.

[12] Jenke, R., Peer, A., & Buss, M. (2013, May). Effect-size-based electrode and feature selection for emotion recognition from EEG. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 1217-1221). IEEE.

[13] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

[14] Noh, Y. K., & Min, B. K. (2014, February). Feature selection for brain-computer interface using nearest neighbor information. In Brain-Computer Interface (BCI), 2014 International Winter Workshop on (pp. 1-3). IEEE.

[15] Akilandeswari, K.,Nasira, G.M., "Bagging of EEG Signals for Brain Computer Interface" , IEEE Xplore,World Congress on Computing and Communication Technologies (WCCCT), 2014 , pp.71, 75 , doi: 10.1109/WCCCT.2014.42.

[16] Akilandeswari, K.,Nasira, "Swarm Optimized Feature Selection of EEG Signals for Brain- Computer-Interface", International Journal of Computational Intelligence and Informatics, Vol. 4: No. 2, October - December 2014.