Weight Optimization of Multilayer Perceptron Neural Network using Hybrid PSO for Improved Brain Computer Interface Data Classification

K. Akilandeswari
Department of Computer Science
Government Arts College (Autonomous)
Salem-7
akilamanivelan@gmail.com

R. Uma Rani
Department of Computer Science
Sri Sarada College for Women (Autonomous)
Salem - 16
umainweb@gmail.com

Abstract- **A Brain-Computer Interface (BCI), also known as Brain Machine Interface (BMI), is a communication system that lets the users to interact with electronic devices by means of control signals acquired from Electroencephalographic (EEG) activity without engaging peripheral nerves and muscles. The preliminary motivation for BCI research was to develop assistive devices for people with locked-in disabilities. Nowadays, researchers are exploring BCI as a novel anthropomorphic interaction channel for daily applications such as robotics, virtual reality, and games. This paper investigates the effect of weight optimization using Hybrid Particle Swarm Optimization (PSO) for Multilayer Perceptron Neural Network classifier which uses features selected by Principal Component Analysis and Hybrid PSO.**

Keywords -Brain Computer Interface (BCI), Electroencephalographic (EEG), Feature Selection, Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), Weight optimization, Multilayer Perceptron Neural Network (MLPNN).

1. INTRODUCTION

Brain-Computer Interface (BCI) is a hardware and software communications system that permits cerebral activity alone to control computers or external devices. The immediate goal of BCI research is to provide communications capabilities to severely disabled people who are totally paralyzed or 'locked in' by neurological neuromuscular disorders, such as amyotrophic lateral sclerosis, brain stem stroke, or spinal cord injury (Nicolas-Alonso L. F., 2012). BCI activates electronic or mechanical devices with brain activity alone. BCIs allow direct brain communication in completely paralyzed patients and restoration of movement in paralyzed limbs through the transmission of brain signals to the muscles or to external prosthetic devices (Birbaumer, 2006).

There are many phases in Brain Computer Interfacing such as (Rao T. K., 2012):

- Signal Acquisition
- Signal Pre-Processing
- Signal Classification
- Computer Interaction

Electroencephalograph (EEG) is an instrument used for recording the electrical activity of the brain. EEG is the variation of the electrical fields in the cortex or on the surface of scalp caused by the physiological activities of the brain. EEG is currently the most widely adopted method for assessing brain activities. Detecting the changes of these waves is critical for understanding of brain function. In clinical applications, spontaneous EEG signals can be divided into several rhythms according to their frequency (Ahmed S. A., 2012). EEG is a graphic representation of the difference in voltage between two different cerebral locations plotted over time. The scalp EEG signal generated by cerebral neurons is modified by electrical conductive properties of the tissues between the electrical source and the recording electrode on the scalp, conductive properties of the electrode itself, as well as the orientation of the cortical generator to the recording electrode. The EEG can be obtained because of the process of current flow through the tissues between the electrical generator and the recording electrode, which is called volume conduction. EEG provides a two-dimensional projection of three-dimensional reality, which means that theoretically it is impossible to determine the location of the EEG generator based on scalp-recorded EEG information alone (Olejniczak, 2006).

There are five categories that cover the most used algorithms in BCI classification systems, and they are: linear classifiers, nonlinear Bayesian classifiers, nearest neighbor classifiers, neural networks, and a combination of classifiers (Lotte F., 2007). In all categories there have been achieved good BCI results, except for the nearest neighbor classifiers, which seems to not handle dimensionality very well. However, neural networks are the most popular in BCI research (Larsen, 2011).

Linear classifiers use the linear functions to classify signals into classes. The most frequently used linear classifiers are Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) (Lotte F., 2007). LDA creates models of the Probability density function respectively. LDA is simple to use and has very low computational requirements. It provides good results. For non-Gaussian distributions LDA may not preserve the complex structure in the data. LDA fails if the discriminatory function is not in mean but in the variance of the data (Senthilmurugan, 2011).

SVM: SVM is a linear classifier that is used by most of the BCI applications. SVM was developed by Vapnik and was driven by statistical learning theory following the principle of structural risk minimization (Ghanbari A. A., 2012). SVM finds a hyper plane to separate the data sets. It separates data sets with clear gap that is as wide as possible to classify them into their relevant category. The hyper plane maximizes the margin that is the distance between the hyper plane and the nearest points from each class that are called as support vectors (Barachant, 2012).

ANNs are non linear classifiers composed of large number of interconnected elements called neurons. Each neuron in ANN simulates the biological neuron and is capable of performing simple computational tasks. The most frequently used neural network is the Multi Layer Perceptron Neural Network (MLPNN) in which, the network is arranged into three layers viz., input layer, hidden layer and output layer. The advantage of MLPNN is that its fast operation, ease of implementation and requiring small training sets (Hekim, 2012).

Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. It is the most important step that affects the performance of a pattern recognition system. Feature selection can serve as a pre-processing tool of great importance before solving the classification problems. The purpose of the feature selection is to reduce the maximum number of irrelevant features while maintaining acceptable classification accuracy. Feature selection is of considerable importance in pattern classification, data analysis, multimedia information retrieval, biometrics, remote sensing, computer vision, medical data processing, machine learning, and data mining applications.

In BCI the feature selection methods are Gain Ranking (IG), Correlation-Based Feature Selection (CFS), ReliefF, Consistency-Based Feature Selection (Consistency) and 1R Ranking (1RR). IG is a very popular and successful feature selection method for high dimensional data, widely used in the area of text classification (Yang, 1997).CFS is a simple and fast feature subset selection method developed by Hall (Hall, 1999). It searches for the “best” subset of features where “best” is defined by a heuristic which takes into consideration two criteria: 1) how good the individual features are at predicting the class and 2) how much they correlate with the other features. Good subsets of features contain features that are highly correlated with the class and uncorrelated with each other.

Relief (Kira, 1992) is an instance-based feature ranking method for two-class problems. ReliefF (Kononenko, 1994) is an extension of Relief for multiclass problems. Relief ranks the features based on how well they distinguish between instances that are near to each other. Among the various methods proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method (Ghanbari A. A., 2012) and Ant Colony Optimization (ACO) based method have attracted a lot of attention (Rashidy Kanan, 2007). These methods attempt to achieve better solutions by using knowledge from previous iterations with no prior knowledge of features.

This paper presents BCI data classification using Bagging with Naïve Bayes classifier and MLPNN and proposed MLPNN weight optimization using Hybrid PSO. The rest of the paper is organized as follows: Section 2 reviews some of the related works available in the literature, section 3 explains the techniques used in this investigation, section 4 presents the results and discussion and section 5 concludes the paper.

2. LITERATURE REVIEW

Zhiping et al., (Zhiping, 2010) proposed a new feature selection method based on PSO for EEG-based Motor-Imagery (MI) BCI systems. The method includes the following two steps: (1) an optimization algorithm, i.e. PSO is used to select the EEG features and classifier parameters; and (2) a voting mechanism is introduced to remove the features redundant, which produced by optimization algorithm. It also compared the proposed method with the GA method. Experiment on single-trial MI EEG classification showed the effectiveness of the proposed method.

Guerrero-Mosquera et al., (Guerrero-Mosquera, 2010) compared the three subsets of features obtained by tracks extraction method that are wavelet transform and fractional Fourier transform. It also compared the performance of each subset in classification tasks using support vector machines and then selects possible combination of features by feature selection methods based on forward–backward procedure and mutual information. Results confirm that fractional Fourier transform coefficients presented very good performance and also the possibility of using some combination of this features to improve the performance of the classifier.

Chum et al., (Chum, 2012) proposed the optimal feature extracting method from the basic power density of EEG signal. In simulation used the dataset from BCI competition III, IV and the data experimented in laboratory. To ensure the improvement of this proposed feature extraction method, it applied the extracted feature into the support vector machine.

LaRocco et al., (LaRocco, 2014) feature reduction and classifier structures were investigated. This paper presented a single feature corresponding to the maximum of average distance between events and non-events (ADEN) on unbalanced data yielded a phi correlation of 0.94 on the mock data with an SNR of 0.3, compared with a phi coefficient of 0.00 for PCA. This simulation has demonstrated strong potential compared to other feature selection/reduction methods.

Kimovski et al., (Kimovski, 2014) proposed master-worker implementations of two different parallel evolutionary models, the parallel computation of the cost functions for the individuals in the population, and the parallel execution of evolutionary multi-objective procedures on subpopulations. It showed the benefits of parallel processing not only for decreasing the running time, but also for improving the solution quality.

Nasehi & Pourghassem (Nasehi, 2011) presented a novel effective feature selection based on Statistical-Principal Component Analysis (S-PCA) and wavelet transform (WT) features in medical and BCI application. In this algorithm, S-PCA is used to select ten effective features from among WT features. It use KNN classifier and seven different signals of brain activities to evaluate the proposed method. The results indicated the improvement of the classification performance in comparison with current methods.

Jenke et al., (Jenke, 2013) proposed a statistically-motivated electrode/feature selection procedure, based on Cohen's effect size f2. This paper compared inter- and intra-individual selection on a self-recorded database. Classification is evaluated using Quadratic Discriminant Analysis (QDA). While highest accuracies up to 57,5% (5 classes) are reached by applying intra-individual selection, inter-individual analysis successfully finds features that performed with lower variance in recognition rates across subjects than combinations of electrodes/features suggested in literature.

Rodríguez - Bermúdez et al., (Rodríguez-Bermúdez, 2013) presented an efficient embedded approach for feature selection and linear discrimination of EEG signals. The proposed method efficiently selects and combines the most useful features for classification with less computational requirements. Least Angle Regression (LARS) is used for properly ranking each feature and, then, an efficient Leave-One-Out (LOO) estimation based on the PRESS statistic is used to choose the most relevant features. Experimental results on motor-imagery BCIs problems are provided to illustrate the competitive performance of the proposed approach against other conventional methods.

Gan et al., (Gan, 2011) proposed a filter-dominating hybrid Sequential Forward Floating Search (SFFS) method, aiming at high efficiency and insignificant accuracy sacrifice for high-dimensional feature subset selection. Experiments with this new hybrid approach have been conducted on BCI feature data, in which both linear and nonlinear classifiers as wrappers and Davies-Bouldin index and mutual information based index as filters are alternatively used to evaluate potential feature subsets. Experimental results have demonstrated the advantages and usefulness of the proposed method in high-dimensional feature subset selection for BCI design.

Yu et al., (Yu, 2013) proposed a discriminative feature extraction algorithm based on power bands with PCA. The raw EEG signals from the motor cortex area were filtered using a bandpass filter with μ and β bands. This research considered the power bands within a 0.4 second epoch to select the optimal feature space region. Also, the total feature dimensions were reduced by PCA and transformed into a final feature vector set. The selected features were classified by applying a Support Vector Machine (SVM). The proposed method was compared with a state-of-art power band feature and shown to improve classification accuracy.

A BCI collects data from sensors, and the data are discriminated using information in a high-dimensional space. Noh et al., (Noh, 2014) showed how the nearest neighbor method can be exploited by properly trimming the non-informative direction for a distance calculation, and estimate the Jensen-Shannon divergence more accurately. Through experiments with synthetic data, it showed how the proposed method outperforms a conventional nearest neighbor method as well as other feature selection methods with a large margin.

Kołodziej et al., (Kołodziej, 2011) analyzed the EEG signal and translate patient intentions for simple commands. Signal processing methods are a very important step in such systems. Signal processing includes: preprocessing, feature extraction, feature selection and classification This paper presented the results of implementing Linear Discriminant Analysis as a feature reduction tool.

Lan (Lan, 2011) proposed that feature manipulations, including feature extraction, feature selection and dimensionality reduction, can solve or at least partly solve the robustness, real-time and non-stationary problems. This research focused on two BCI applications: Augmented Cognition (AugCog) and single trial ERP detection. Experimental results showed that the proposed methods improve the performance of BCI systems compared with these baseline systems.

Bhattacharyya et al., (Bhattacharyya, 2014) proposed an efficient feature selection technique, realized by means of an evolutionary algorithm, which attempts to overcome some of the shortcomings of several state-of-the-art approaches in this field. Also presented an efficient memetic algorithm for the optimization purpose. Extensive experimental validations have been conducted on two real-world datasets to establish the efficacy of our approach. These results compared to existing algorithms and have established the superiority of this algorithm to the rest.

Koprinska (Koprinska, 2010) evaluated feature selection methods for classification of BCI data. The methods tested with ten classification algorithms, representing different learning paradigms, on a benchmark BCI competition dataset. The results showed that all feature selectors significantly reduced the number of features and also improved accuracy when used with suitable classification algorithms.

3. METHODOLOGY

Data Set I from BCI Competition III is used for performing the proposed work. Features extraction is performed using WHT and bagging with different classifiers are compared (Akilandeswari K, 2014). The extracted features are used for feature selection based on PSO (Akilandeswari K, 2014) and Hybrid PSO (Akilandeswari K, 2015). The selected features are classified using Bagging with different classifiers and MLPNN. To improve the classification accuracy, weight optimization using Hybrid PSO for MLPNN is proposed.

3.1. Dataset

BCI Competitions are organized in order to foster the development of improved BCI technology by providing an unbiased validation of a variety of data analysis techniques. In each competition a variety of data sets was made publicly available in a documented format via internet (Blankertz B., 2003), (Blankertz B., 2005), (Blankertz B. M., 2008). Each data set is a record of brain signals from BCI experiments of leading laboratories in BCI technology split into two parts: one part of labeled data ('training set') and another part of unlabeled data ('test set'). Researchers worldwide could tune their methods to the training data and submit the output of their translation algorithms for the test data. The truth about the test data was kept secret until, after the deadline, it was used to evaluate the submissions. This procedure guarantees that the assessment of performance is not biased by overfitting the selection of methods and the choice of their parameters to the data.

The dataset used to evaluating the proposed method is Data Set I from BCI Competition III. In BCI experiments, a subject performs imagined movements of left small finger or tongue. 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.

3.2. Hybrid PSO

The Hybrid PSO algorithm starts with an initial swarm of K particles. Each particle vector corresponds to a candidate solution of the underlying problem. Then, all of the particles repeatedly move until a maximal number of iterations have been passed. During each iteration, the particle individual best and swarm's best positions are determined. The particle adjusts its position based on the individual experience (pbest) and the swarm's intelligence (gbest). To expedite the convergence speed, all of the particles are further updated using the hill-climbing heuristic before entering the next iteration.

General structure of hybrid PSO

```
Begin
Create and initialize:
While(stop condition is false)
Begin
        Evaluation
        Update velocity and position
        Mutation
End
End
```

3.3. Multi Layer Perceptron Neural Networks (MLPNN)

Multi-layered Perceptron (MLP) is used for the neural networks with a structure of input layer, one or more hidden layers and an output layer. Each of the layers consists of inter-connected assembly of simple processing elements called neurons. These processing elements are organized in a layered fashion. Each neuron in a layer is connected to the neuron in the subsequent layer and so on. The interconnections between layers are called weights. Despite of their simplified structure, neural networks have ability to mimic human characteristics of problem solving via learning and generalization. MLP can be used to model non-linear systems due to their ability to learn the system behavior under inspection from samples (Hamidi, 2012).

MLPNN neural networks are trained using the Back Propagation (BP) algorithm which is a gradient based supervised learning method. According to the algorithm, a mean squared error between the predicted and target values for the given input parameters is propagated backward to adjust the interconnection between neurons in order to minimize the pre-defined error. In this structure each neuron in a layer is mapping the sum-of weighted input into an activation level that is determined by an activation function. The most commonly used activation functions are the sigmoid, the tangent hyperbolic, and the linear activation function.

3.4. Proposed Weight Optimization of MLPNN using PSO and Hybrid PSO

The serious constraint imposed for the usage of BP algorithm is that the hidden layer neuron function should be differentiable. If the inputs and desired outputs of a function are known then BP can be used to determine weights of the neural network by minimizing the error over a number of iterations. The weight update equations of all the layers (input, hidden, output) in the MLPNN are almost similar, except that they differ in the way the local error for each neuron is computed. The error for the output layer is the difference between the desired output (target) and actual output of the neural network. Similarly, the errors for the neurons in the hidden layer are the difference between their desired outputs and their actual outputs.

In a MLPNN, the desired outputs of the neurons in the hidden layer cannot be known and hence the error of the output layer is back propagated and sensitivities of the neurons in the hidden layers are calculated. The learning rate is an important factor in the BP. If it is too low, the network learns very slowly and if it is too high, then the weights and the objective function will diverge. So an optimum value should be chosen to ensure global convergence which tends to be difficult task to achieve. A variable learning rate will do better if there are many local and global optima for the objective function (Gudise, 2003).

In this paper, the PSO and hybrid PSO are used of weight optimization of the MLPNN. For training a neural network using the PSO, the fitness value of each particle (member) of the swarm is the value of the error function evaluated at the current position of the particle and position vector of the particle corresponds to the weight matrix of the network. The particle stores the minimum error encountered by the particles.

The steps of the method.

1. The weights and biases of the MLPNN are initialized randomly.
2. The network is then trained using the particles initial positions (weights and biases).
3. Then, the feed-forward neural network will produce the learning error (particles fitness) based on
4. initial weight and bias.
5. The learning error (mean square error) at current epoch will be reduced by changing the particles position, which will update the weight and bias of the network. The "pbest" value (each particle‟s lowest learning error so far) and "gbest" value (lowest learning error found in entire learning process so far) are applied to the velocity update equation to produce a value for positions adjustment to the best solution. Then the new sets of positions (weights and biases) are produced by adding the calculated velocity value to the current position value.
6. This process is repeated until the maximum numbers of iteration are met.

For training a neural network using the Hybrid PSO, the particles are further updated using the hill-climbing heuristic before entering the next iteration as mentioned in the general structure of Hybrid PSO,.

4. RESULTS AND DISCUSSION

The experiments are conducted using Data Set I from BCI Competition III. The Mat lab is used for feature extraction and feature selection and WEKA tool is used for classification. The features are extracted using WHT. PSO and hybrid PSO feature selection methods are combined with PCA for obtaining optimized feature subset. The selected features are classified using Naïve Bayes, MLPNN and proposed PSO-MLPNN and hybrid PSO-MLPNN. The WHT with PCA-PSO-MLP-hybrid PSO improved the classification accuracy by 2.7625% when compared to WHT with PCA-bagging with NB tree. The WHT with PCA-PSO-MLP-hybrid PSO improved the classification accuracy by 1.8506% when compared to WHT with PCA-hybrid PSO-bagging with NB.

Table: 1 Comparison of Classification Accuracy with various techniques

Techniques	Classification Accuracy
WHT with PCA – Bagging with NB tree	94.96
WHT with PCA and PSO - Bagging with NB	95.68
WHT with PCA and Hybrid PSO - Bagging with NB	95.83
WHT with PCA and PSO - MLP	96.43
WHT with PCA and Hybrid PSO - MLP	97.02
WHT with PCA and PSO - MLP - Hybrid PSO	97.62

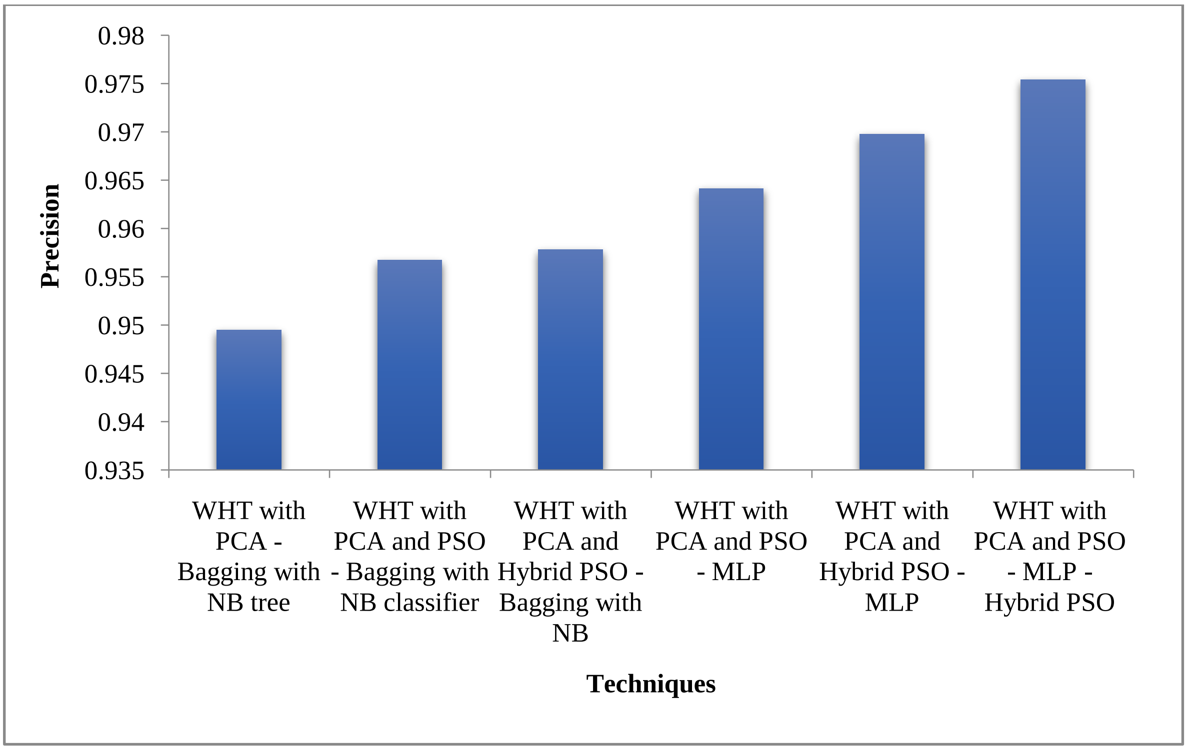

Figure 1. Precision obtained with different techniques

The WHT with PCA-PSO-MLP-hybrid PSO improved the precision by 2.7413% when compared to WHT with PCA-bagging with NB tree. The WHT with PCA-PSO-MLP-hybrid PSO improved the precision by 1.9091% when compared to WHT with PCA-hybrid PSO-bagging with NB.

The WHT with PCA-PSO-MLP-hybrid PSO improved the recall by 2.6924% when compared to WHT with PCA-bagging with NB tree. The WHT with PCA-PSO-MLP-hybrid PSO improved the recall by 1.8172% when compared to WHT with PCA-hybrid PSO-bagging with NB.

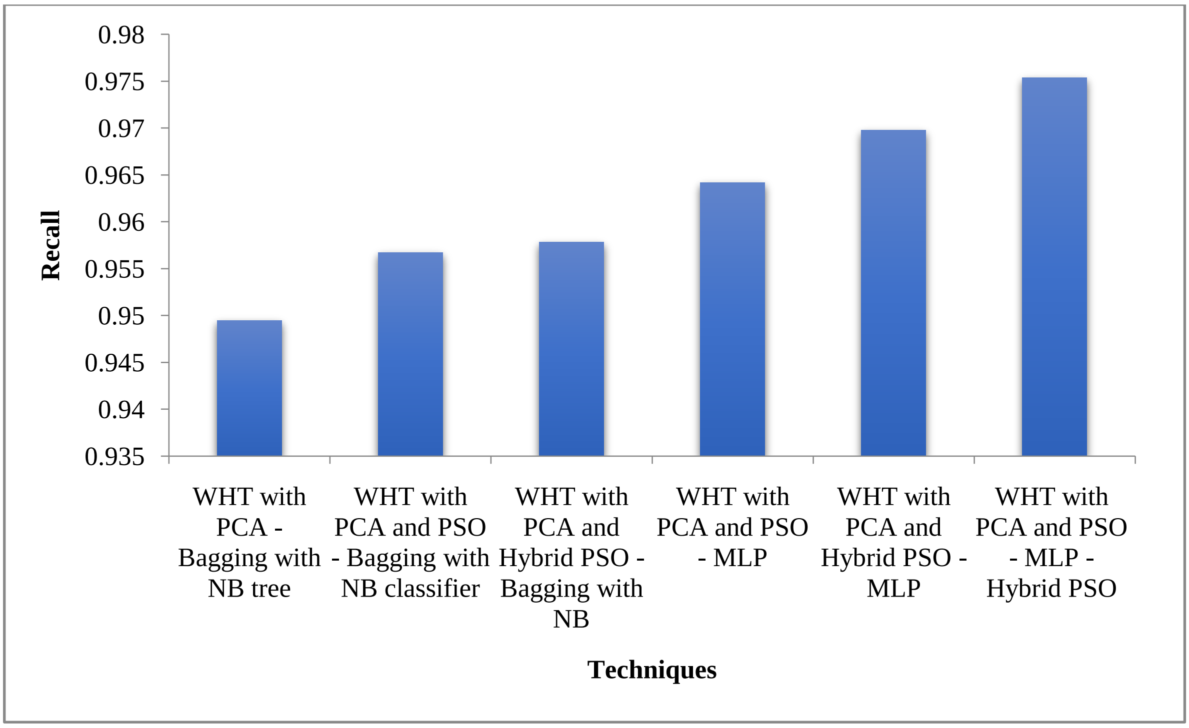

Figure 2. Recall obtained with different techniques

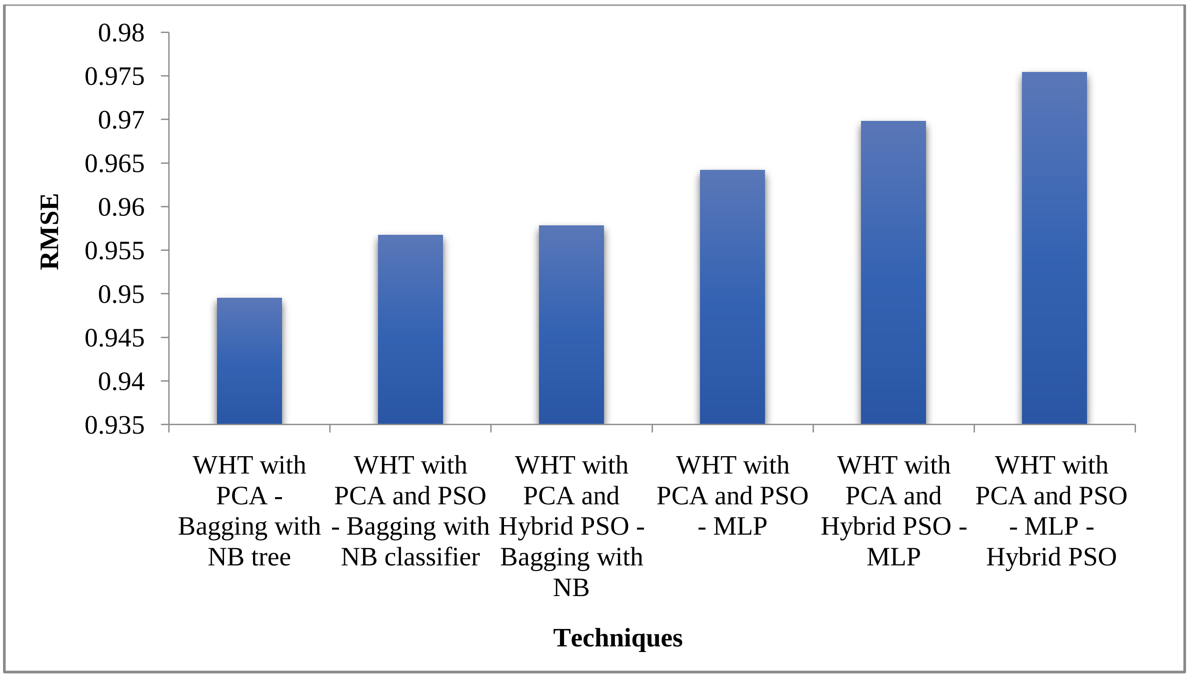

Figure 3. Root Mean Square Error obtained with different techniques

The WHT with PCA-PSO-MLP-hybrid PSO reduced Root Mean Square Error (RMSE) by 15.493% when compared to WHT with PCA-bagging with NB tree. The WHT with PCA-PSO-MLP-hybrid PSO reduced RMSE by 11.0402% when compared to WHT with PCA-hybrid PSO-bagging with NB.

5. CONCLUSION

As computerized systems are becoming one of the main tools for making people's lives easier and with the ongoing growth in the BCI field, it is becoming more important to understand brain waves and analyze EEG signals. Experiments were conducted with Data Set I from BCI Competition III. Features extracted by WHT are used for feature selection based on PSO and hybrid PSO. MLPNN for classification is experimented with BP training. The problem of the BP algorithm is that it is very often trapped in local minima and the learning and adaptation speed are very slow. To improve the efficacy of the MLPNN for classification, a PSO and hybrid PSO for weight optimization is proposed. Results demonstrate that weight optimization of the MLPNN using Hybrid PSO significantly improves the classification accuracy.

REFERENCES

Ahmed S. A., R. D. (2012). Alpha Activity in EEG and Intelligence. International Journal of Advanced Information Technology (IJAIT) , 12 (1).

Akilandeswari K, N. G. (2015). A Hybrid Meta Heuristic Feature Selection Algorithm for Brain Computer Interface. International Journal of Science and Engineering Research , 6 (5), 910-916.

Akilandeswari K, N. G. (2014). Bagging of EEG Signals for Brain Computer Interface. World Congress on Computing and Communication Technologies (WCCCT) (pp. 71-75). Trichy: IEEE Xplore.

Akilandeswari K, N. G. (2014). Swarm Optimized Feature Selection of EEG Signals for Brain Computer Interface. International Journal of Computational Intelligence and Informatic, 4.

Barachant, A. B. (2012). BCI Signal Classification using a Riemannian-based kernel. 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.

Bhattacharyya, S. S. (2014). Automatic feature selection of motor imagery EEG signals using differential evolution and learning automata. Medical & biological engineering & computing .

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

Blankertz, B. (2003). BCI competition. htm:/fida. ftrst. rag. de/projects/bci/compeu'tion (web page).

Blankertz, B. (2005). BCI competition III webpage. ht tp://ida. first. fraunhofer. de/projects/bci/competition iii.

Blankertz, B. M. (2008). The BCI Competition III.

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

Chum, P. P. (2012). Optimal EEG Feature Selection by Genetic Algorithm for Classification of Imagination of Hand Movement. 38th Annual Conference on IEEE Industrial Electronics Society IECON. IEEE 1561-1566.

Gan, J. Q. (2011). A hybrid approach to feature subset selection for brain-computer interface design. Springer Berlin Heidelberg. Intelligent Data Engineering and Automated Learning-IDEAL, 279-286.

Ghanbari, A. A. (2012). Brain computer interface with genetic algorithm. International Journal of Information , 2 (1).

Ghanbari, A. A. (2012). Brain Computer Interface with Wavelets and Genetic Algorithms.

Gudise, V. G. (2003). Comparison of particle swarm optimization and backpropagation as training algorithms for neural network. Swarm Intelligence Symposium. IEEE, 110-117.

Guerrero-Mosquera, C. V. (2010). EEG feature selection using mutual information and support vector machine: A comparative analysis. Annual International Conference In Engineering in Medicine and Biology Society (EMBC), 2010. IEEE, 4946-4949.

Hall, M. A. (1999). Feature selection for discrete and numeric class machine learning.

Hamidi, J. (2012). Application of Multi-Layered Perceptron Neural network (MLPNN) to Combined Economic and Emission Dispatch.,. International Journal of Computer and Electrical Engineering , 242-246.

Hekim, M. (2012). ANN-based classification of EEG signals using the average power based on rectangle approximation window. Electrical Review , 88.

Jenke, R. P. (2013). Effect-size-based electrode and feature selection for emotion recognition from EEG. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 1217-1221.

Kimovski, D. O. (2014). Feature selection in high-dimensional EEG data by parallel multi-objective optimization. IEEE International Conference on Cluster Computing. IEEE, 314-322.

Kira, K. &. (1992). A practical approach to feature selection. ninth international workshop on Machine learning. Morgan Kaufmann Publishers Inc, 249-256.

Kołodziej, M. M. (2011). Implementation of linear discriminant analysis as a feature selection technique of EEG signal for brain-computer interface. International Conference Communication Software and Networks (ICCSN), 2011.

Kononenko, I. (1994). Estimating attributes: analysis and extensions of RELIEF. Machine Learning:ECML (94), 171-182.

Koprinska, I. (2010). Feature selection for brain-computer interfaces. In New Frontiers in Applied Data Mining. Springer Berlin Heidelberg., 106-117.

Lan, T. (2011). Feature extraction feature selection and dimensionality reduction techniques for brain computer interface. International Conference on Intelligent Systems Design and Applications (ISDA).

LaRocco, J. I. (2014). Optimal EEG feature selection from average distance between events and non-events. 36th Annual International Conference of the IEEE in Engineering in Medicine and Biology Society (EMBC). IEEE, 2641-2644.

Larsen, E. A. (2011). Classification of EEG Signals in a Brain-Computer Interface System. international workshop on Machine learning. IEEE.

Lotte F., C. M. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. Journal of neural engineering , 4.

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

Nicolas-Alonso L. F., &. G.-G. (2012). Brain computer interfaces, a review. Sensors , 12(2), 1211-1279.

Noh, Y. K. (2014). Feature selection for brain-computer interface using nearest neighbor information. International Winter Workshop on Brain-Computer Interface (BCI). IEEE, 1-3.

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

Rao T. K., L. M. (2012). An exploration on brain computer interface and its recent trends. arXiv preprint arXiv:1211.2737.

Rashidy Kanan, H. F. (2007). Face recognition system using ant colony optimization-based selected features. CISDA 2007. IEEE Symposium (p. 5). InComputational Intelligence in Security and Defense Applications.

Rodríguez-Bermúdez, G. G.-L.-G.-D. (2013). Efficient feature selection and linear discrimination of EEG signals. Neurocomputing , 161-165.

Senthilmurugan, M. L. (2011). Classification in EEG-based Brain Computer Interfaces Using Inverse Model. International Journal of Computer Theory and Engineering , 3(2), 274-276.

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

Yang, Y. &. (1997). A comparative study on feature selection in text categorization. ICML , 97, 412-420.

Yu, X. P. (2013). Discriminative Power Feature Selection Method for Motor Imagery EEG Classification in Brain Computer Interface Systems. International Journal of Fuzzy Logic and Intelligent Systems (13(1)), 11-17.

Zhiping, H. G. (2010). A new EEG feature selection method for self-paced brain-computer interface. Intelligent Systems Design and Applications (ISDA). IEEE, 845-849.