International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:6, 2015

Multi-Layer Perceptron Neural Network Classifier with Binary Particle Swarm Optimization Based Feature Selection for Brain-Computer Interfaces

K. Akilandeswari, G. M. Nasira

***Abstract*—**Brain-Computer Interfaces (BCIs) measure brain signals activity, intentionally and unintentionally induced by users, and provides a communication channel without depending on the brain's normal peripheral nerves and muscles output pathway. Feature Selection (FS) is a global optimization machine learning problem that reduces features, removes irrelevant and noisy data resulting in acceptable recognition accuracy. It is a vital step affecting pattern recognition system performance. This study presents a new Binary Particle Swarm Optimization (BPSO) based feature selection algorithm. Multi-layer Perceptron Neural Network (MLPNN) classifier with backpropagation training algorithm and Levenberg-Marquardt training algorithm classify selected features.

***Keywords*—**Brain-Computer Interfaces (BCI), Feature Selection (FS), Walsh–Hadamard Transform (WHT), Binary Particle Swarm Optimization (BPSO), Multi-Layer Perceptron (MLP), Levenberg–Marquardt algorithm.

I. Introduction

BCI is a control and communication system not depending on the brain's normal neuromuscular output channels. User's intent is conveyed by brain signals (like electroencephalography (EEG)) instead of peripheral nerves and muscles. These brain signals do not depend on neuromuscular activity for their generation [1]. A BCI, also called a brain machine interface or direct neural interface, is a straight communication pathway between a human or brain cell culture or animal brain and an external device.

In a BCI, computers accept commands from a brain or send it signals, but not both. Exchanging information both ways i.e. between brains and external devices, are yet to be successfully implanted in either humans or animals [2]. BCIs, which focus on motor Neuro-prosthetics, aim to restore movement in paralyzed individuals or devices to assist them like interfaces with computers or robotic arms.

Feature Selection (FS) finds relevant components for which a classifier's performance is the best. So, feature selection and induction, i.e., process of learning appropriate classifiers are related. Depending on how both algorithms are related, three approaches needed are:

1. Embedded methods: induction and FS algorithms are indivisible;
2. Filter methods: FS precedes induction algorithm; and
3. Wrapper methods: FS algorithm uses induction algorithm. [3].

Filter based FS methods are faster than wrapper based methods as they depend on some estimation of individual features or features subset importance [4]. Compared to filter methods, wrapper based methods are accurate as feature subsets importance is measured by a classification algorithm.

FS selects a subset of original features according to some criteria and is a frequently used dimensionality data mining reduction technique. It reduces features, removes irrelevant and redundant data, ensuring immediate effect for applications: speeding up data mining algorithm, and improving mining performance like predictive accuracy and result comprehensibility [5].

FS reduces data dimensionality which means faster classifier building and producing more compact and easier to interpret classification rules. FS methods are categorized on what they evaluate and rank: individual features or features subsets. Correlation-based Feature Selection (CFS) and consistency evaluate features subset and produce one feature subset; Information Gain (IG), Relief and 1R Ranking (1RR) evaluate features individually and rank them; a feature subset selection is achieved by selecting highest N ranked features or features with a value above t, where N and t are user-specified thresholds [6].

Over the last decade, electrical recordings from the brain's surface Electrocorticography (ECoG) has been recognized as a promising signal platform for BCI research. ECoG is acquired by placing electrodes below the skull, either above (epidural) or below (subdural) dural matter, but not within a brain's parenchyma. Compared to signals from scalp EEG and intra-parenchymal single neuronal recordings, ECoG recording characteristics suit them for neuroscience research and translational opportunities. These include high spatial resolution, signal fidelity, noise resistance, and substantial robustness over long recording periods. ECoG recordings balance fidelity and clinical practicality [7]. An intermediate BCI methodology with ECoG activity from the cortical surface can be a practical alternative to such extremes [8].

ECoG records electrical signals from the brain's surface in patients before surgery. ECoG is less invasive than neuronal recordings as the brain is not penetrated and so has a higher Signal-To-Noise Ratio (SNR) than EEG, as also higher spectral and spatial resolution. The higher resolution needs re-engineering of signal processing and classification techniques in conventional EEG-based BCIs [9]. Another ECoG issue is

K.Akilandeswari is Research Scholar in Periyar University, Salem, Tamilnadu, India (e-mail: akilaphdscholar2015@gmail.com).
G. M. Nasira is with Chikanna Government Arts College, Tirupur, Tamilnadu, India (e-mail: nasiragm99@yahoo.com).

that it is an invasive procedure performed only for medical needs. So, access to ECoG data and subjects is limited.

The collective behavior of self-organized, decentralized, natural or artificial systems is Swarm Intelligence (SI) consisting of a population of simple agents interacting with one another locally and with their environment [10]. SI studies systems collective behavior composed of many individuals interacting with each other and with the environment. Some significant SI techniques are i) Particle Swarm Optimization (PSO); ii) Ant Colony Optimization (ACO); iii) Artificial Bee Colony Optimization (ABC); and iv) Consultant-Guided Search (CGS).

PSO incorporates swarming behavior of bird flocks, fish schools or bee swarms and even human social behavior from which the idea started. ACO deals with artificial systems inspired from real ants foraging behavior which solved discrete optimization problems [11]. The main idea is indirect communication between ants by chemical pheromone trials, enabling them to find shortest paths between food and their nest.

Artificial Neural Network (ANN) algorithms classify Regions of Interest (ROI) using a method similar to the human brain like learning, understanding, solving problems, and taking decisions. An NN in its general form is a machine designed to model on how the brain performs a specific task or function of interest [32]. ANN architecture has 3 units. The first is the input layer, and its nodes are determined by input parameters. The last layer is an output layer whose nodes are given by the desired output. The layer(s) between input and output layers is the hidden layer(s).

This study presents a BPSO based new feature selection algorithm and Multi-layer Perceptron Neural Network (MLPNN) classifier with back propagation training algorithm and Levenberg-Marquardt training algorithm to classify selected features. The rest of the paper is summarized as follow: Section II discusses related work. Section III explains the methodology. Section IV discusses the results, and Section V concludes the paper.

II. Related Works

Smart Multi-Objective PSO using Decomposition (SDMOPSO) introduced by Moubayed [12] used a decomposition approach presented in Multi-Objective Evolutionary Algorithms based on Decomposition (MOEA/D), through which a multi-objective problem was represented as many scalar aggregation problems which in turn were viewed as swarm particles; each assigning weights to all optimisation objectives. The problem was solved as a Multi-Objective PSO (MOPSO), where every particle used information from a defined neighbours set. SDMOSPO covers binary problems and applies the new binary method to BCI's channel selection issues.

The performance of a high density 348 channels Near-Infrared Spectroscopy (NIRS)-based BCI on 8 healthy subjects when solving arithmetic problems with two difficulty levels mentally and the rest condition was investigated by [13]. A novel method to extract effective features from high density single-trial NIRS data was by using common average reference spatial filtering. The proposed feature extraction method's performance was presented using 5×5-fold cross-validations on one trial NIRS data collected using mutual information-based FS and SVM classifier. Results yielded an average accuracy of 73% and 92% in classifying hard versus easy tasks and hard versus rest tasks respectively using the new method, compared to 46% and 62% respectively using the current method. Results proved the effectiveness of the new method in high density NIRS-based BCI for assessed numerical cognition.

A large number of features were extracted from raw EEG data after which FS and classification were performed by [14], for BCI applications using motor imaginary movements. The minimum Redundancy Maximum Relevance (mRMR) feature selection method which selects relevant and non-redundant feature set quickly was chosen. Many different classifiers proved that FS helped classification performance, and higher classification accuracy was achieved using less features. BCI Competition 2003 3A data set was used for experiments.

An efficient subject-independent procedure for EEG-based BCIs was presented by [15]. Three feature extraction methods including Autoregressive modelling, Wavelet transform, and Power spectral density were applied; then, a new method based on GA wrapped Self Organization Map (SOM) feature selection selected most related features using leave-one-subject-out cross-validation strategy. Experiment results proved that the proposed algorithm based on GA wrapped SOM feature selection is efficient to design subject-independent BCIs and can distinguish different cognitive tasks of various individuals effectively.

A new PSO based FS method EEG-based Motor-Imagery (MI) Self-paced BCI (SBCI) systems proposed by [16] included two steps: first, an optimization algorithm, i.e. PSO selected EEG features and classifier parameters; and second a voting mechanism removed redundant features produced by optimization algorithm. The proposed method included a GA. Experiment on single-trial MI EEG classification showed the new method's effectiveness.

New effective FS based on Statistical-Principal Component Analysis (S-PCA) and Wavelet Transform (WT) features in medical and BCI application was proposed by [17]. Signals were sent to six sub-bands by four mother wavelet (sym6, db5, bior1.5, and robio2.8). Then five features (like number of zero coefficients, smallest/largest coefficients, mean and standard deviation of coefficients) were extracted from a sub-band as feature vector. S-PCA selected ten effective features from WT features. KNN classifier and seven different brain activity signals evaluated the new method. Results indicate improved classification performance compared to current methods.

A two-dimensional BCI using event-related de-synchronization and event-related synchronization associated with human natural behaviour so that users did not need long-term training or high mental loads to maintain concentration was proposed by [18]. GA-based Mahalanobis Linear Distance (MLD) classifier and Decision Tree Classifier (DTC) were used for FS and classification, and a model adaptation method was used for better decoding of human movement intention from EEG activity. Results demonstrated good control accuracy for this four-class classification: a high of 77.1% in online control with physical movement.

Various methods exist to classify ECoG signals different in features and classifiers. Used features depend on extracted features, feature reduction methods and feature selection measures. Different algorithms with different results could be used for a specific data set. The best algorithm to do a five-class finger flexion classification to choose flexed finger among a hand's fingers was tried out by [19]. To achieve this, after feature extraction, different feature reduction methods and classification examined training data, and the best algorithm was selected according to results.

An ensemble classifier using PCA features to identify evoked P300 signals from EEG recordings was proposed by [20]. The proposed method's performance was examined with different linear classifiers, on datasets provided by BCI competition III. Results showed 91% classification accuracy with the new method indicating significant improvement in classification accuracy compared to conventional feature extraction and classification approaches. The new method resulted in low across-subjects variability compared to other methods with minimal parameter tuning needed. This could be useful in mobile platform P300 applications.

A novel time-frequency selection method based on a criterion called Time-Frequency Discrimination Factor (TFDF) to extract discriminative Event-Related De-synchronization (ERD) features for BCI data classification was proposed by [21]. Compared to current methods, the new approach generated better classification performance (mean kappa coefficient= 0.62) on experimental data from BCI competition IV dataset II b, with two bipolar channels alone.

A classification-guided (wrapper) method for time-domain NIRS feature extraction to classify left/right hand movements was presented by [22]. NIRS data from two subjects showed that a rank-based wrapper in conjunction with polynomial SVMs achieved 100% sensitivity and specificity separating left/right hand movements (5-fold cross-validation). Results showed potential of wrapper methods to classify NIRS signals for BCI applications.

A method to classify single-trial Event-Related Potentials (ERPs) combining Lifting Wavelet Transform (LWT), SVM, and PSO was proposed by [23]. LWT filters, set of EEG channels and SVM parameters that maximize classification accuracy were searched using PSO. The method's performance was through offline analyses on BCI Competitions II and III datasets. The proposed method achieved similar or higher classification accuracy than that by other methods adapting wavelet basis functions and channel sets matching P300 ERP's time-frequency and spatial properties.

Bhattacharyya et al. [24] tried to reduce a dataset's redundant features to improve classification accuracy. Differential Evolution with Temporal Difference Q-Learning based clustering algorithm reduced features and acquired corresponding accuracy. The new method's superiority was proved by comparing it with three classification methods including LDA, kNN, and SVM-Radial Basis Function. Self-Adaptive Differential Evolution, Differential Evolution/ current-to-best/l, PSO and GA-based clustering approaches were used to study the relative performance of the new adaptive memetic algorithm-based clustering technique regarding runtime and classification accuracy.

For a P300-based BCI, a Mutual Information based Channel Selection (MICS) presented by [25] iteratively chose a new channel with maximal dependency to target class and minimal dependency to earlier selected channels. The new method's evaluation on data set II from the third BCI competition showed that MICS enhanced P300 detection rate compared to other state of art channel selection methods developed for P300 speller BCI.

III. Methodology

This work applied Walsh-Hadamard transform for feature extraction and feature selection using BPSO. Multilayer Perceptron NN with back propagation training algorithm and Levenberg-Marquardt training algorithm was used for feature classification.

A Dataset

Data Set I from BCI Competition III includes motor imagery in ECoG recordings and session-to-session transfer provided by Eberhard-Karls University, Tübingen, Germany, Department of Computer Engineering and Department of Medical Psychology and Behavioural Neurobiology (Niels Birbaumer), the Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany (Bernhard Schökopf), and Universität Bonn, Germany.

A subject performs imagined movements of left small finger or tongue in BCI experiments. Recordings were made at 1000Hz. Recorded potentials were kept as microvolt values after amplification. The trials had an imagined tongue/imagined finger movement recorded for 3 seconds. Recording intervals started 0.5 seconds after visual cue end to prevent data reflecting visually evoked potentials. The dataset has two classes, 64 ECoG channels (0.016-300Hz), 1000Hz sampling rate, and 278 training and 100 test trials.

B Walsh–Hadamard Transform (WHT)

WHT is a non-sinusoidal, orthogonal transformation technique decomposing a signal into basis functions set which are Walsh functions with rectangular and square waves with values of +1 or –1. WHT are also called Walsh or Walsh-Fourier transforms. WHT returns sequency values [26]. Sequency is a generalized frequency notion and defined as one-half of an average number of zero-crossings per unit time interval. To estimate signal frequencies in the original signal, every Walsh function has a unique sequency value.

WHT of a signal x, of size $N = 2^n$, is matrix-vector product $\mathrm{WHT}_{N}x$, (as in (1)) where [27]

$$WHT_N = \bigotimes_{i=1}^{n} DFT_2 = \overbrace{DFT_2 \otimes \ldots \otimes DFT_2}^{n} \quad (1)$$

ISSN: 2517-9942
Vol:9, No:6, 2015

Matrix is a 2-point DFT matrix and $\otimes$denotes tensor or Kronecker product. Tensor product of 2 matrices is obtained by replacing every entry of first matrix by that element multiplied by second matrix.

A fast transform algorithm is a sparse factorization of transform matrix and refers to each factor as a stage. The Fast WHT (FWHT) is utilized to get images local structure. This basis function obtains digital numbers in a sense of coefficients [28]. When the coefficients are normalized by dc coefficient of expansion, i.e., local image's average gray value, then they measure purely local structure independent of modality. These numbers are normalized to get a unique number, which is used as feature for image registration.

C Particle Swarm Optimization (PSO) for Feature Selection

Individuals in a particle swarm follow simple behavior: to emulate success of neighboring individuals and their own successes. The collective behavior that emerges from this is that of discovering optimal regions of high dimensional search space [29]. A PSO algorithm maintains a particle swarm, where every particle represents a potential solution. In analogy with evolutionary computation paradigms, a swarm is similar to a population, and a particle is similar to an individual.

Individuals in PSO are referred to as particles flown through hyper-dimensional search space guided by a leader(s) whose performance affects other population particles speed and direction. Each particle's position changes according to its own experience and that of neighbors [30]. A particle represents a problem solution while its position is a multidimensional vector where every dimension is a problem space variable. A PSO Algorithm [31] is given in Fig. 1.

Input: Randomly initialized position and
velocity of the particles: $X_i(0)$ and $V_i(0)$
Output: Position of the approximate global optima X^*
Begin
While terminating condition is not reached do
Begin
for i = 1 to number of particles
Evaluate the fitness: $=f(X_i)$;
Update p_i and g_i;
Adapt velocity of the particle;
Update the position of the particle;
increase i;
end while
end

Fig. 1 PSO algorithm

In binary PSO, a particle's personal best and global best is updated as in a real-valued version. The difference between binary PSO with real-valued version is that particles velocities are defined in terms of probabilities that a bit changes to one. Using this definition, velocity must be restricted within a range [0,1]. So, an introduced map maps all real valued numbers of velocity to a range [0,1]. Normalization function used is a sigmoid function as in (2):

$$V_{ij}^{'}(t) = sig(V_{ij}(t)) = \frac{1}{1+e^{-V_{ij}(t)}} \tag{2}$$

and the particle's new position is obtained using (3) [39]

$$x_{ij}(t+1) = \begin{cases} 1 \text{ if} & r_{ij} < sig(v_{ij}(t+1)) \\ 0 & otherwise \end{cases} \tag{3}$$

where r_{ij} is a uniform random number in range [0,1]. The algorithm proposed for binary PSO is summarized in Fig. 2 [40].

D Multi-Layer Perceptron (MLP)

MLP is a feed forward ANN model that maps sets of input data to a set of appropriate output. An MLP has many layers of nodes in a directed graph, with each layer connected to the next. Except for input nodes, a node is a neuron (processing element) with a nonlinear activation function [33]. MLP taking advantage of supervised learning technique called back propagation trains the network. MLP is a modification of standard linear perceptron and distinguishes data not linearly separable.

An MLP has many layers of neurons: an input layer, one or more hidden layers, and an output layer. Each neuron's input is connected with output of previous layer's neurons whereas output layer neurons determine input feature vector class [34]. In MLP NN with sigmoidal outputs using logistic function $1/(1+e^{-x})$ has outputs within the range [0,1] or tanh(x) with range [-1,1], the back propagation algorithm requires multiplication by function derivative to change a weight connected to the neuron and so carry out gradient descent [35]. Input layer was assigned a constant weight of 1. A learning rule was applied after output presentation. It used back propagation, a supervised learning method, which calculates mean-squared error between actual/expected outputs. Error value is then propagated backwards in the network and small changes made to weights in every layer [36]. Weight changes are calculated to reduce error signal. The whole process is repeated for every trial, and the cycle reiterated till overall error value drops below a pre-determined threshold.

1. Initialize the swarm X_i, the position of particles are randomly initialized within the hypercube. Elements of X_i are randomly selected from binary values 0 and 1.
2. Evaluate the performance F of each particle, using its current position $X_i(t)$.
3. Compare the performance of each individual to its best performance so far: if $F(X_i(t))<F(P_{ibest})$:
$F(P_{ibest})=F(X_i(t))$
$P_{ibest} = X_i(t)$
4. Compare the performance of each particle to the global best particle: if $F(X_i(t))<F(P_{gbest})$:
$F(P_{gbest})=F(X_i(t))$
$P_{gbest} = X_i(t)$
5. Change the velocity of the particle, $\vec{V}_i^0$ and $\vec{V}_i^1$
6. Calculate the velocity of change of the bits, $\vec{V}_i^c$.
7. Generate the random variable rij in the range: (0,1) . Move each particle to a new position.
8. Go to step 2, and repeat until convergence.

Fig. 2 Binary PSO algorithm

The Levenberg–Marquardt algorithm developed independently by Kenneth Levenberg and Donald Marquardt, gives a numerical solution to minimizing a nonlinear function. It is fast with stable convergence. In ANN, this algorithm suits training small- and medium-sized problems [37]. The basic idea of Levenberg–Marquardt algorithm is to perform a combined training process: around an area with complex curvature, Levenberg–Marquardt algorithm switches to steepest descent algorithm, till local curvature is proper to ensure a quadratic approximation; it then approximately becomes a Gauss–Newton algorithm, which speeds up convergence significantly.

Levenberg – Marquardt algorithm was specifically designed to reduce sum-of-square error functions of form as in (4):

$$E = \frac{1}{2}\sum k(e_k)^2 = \frac{1}{2}\|e\|^2 \tag{4}$$

where e_k is error in k^{th} exemplar or pattern and e is a vector with element e_k. In Levenberg-Marquardt algorithm, error function is minimized, while step size is kept small to ensure linear approximation validity, which is accomplished by using a modified error function of form [38], as in (5):

$$E = \frac{1}{2}\left\| e(j) + \partial e_k / \partial w_i (w_{(j+1)} - w_{(j)}) \right\|^2 + \lambda \left\| w_{(j+1)} - w_{(j)} \right\| \tag{5}$$

where λ is a parameter governing step size. Minimizing modified error regarding $w_{(j+1)}$ gives as in (6):

$$w_{(j+1)} = w_{(j)} - \left(Z^T Z + \lambda I\right)^{-1} Z^T e(j) \tag{6}$$

Very large values of λ amount to standard gradient descent, while very small values λ of amount to a Newton method.

IV. Results and Discussion

Table I shows the parameters used. Tables II-IV and Figs. 3-6 are the Classification Accuracy and RMSE, Precision and Recall for Finger, and Precision and Recall for Tongue respectively.

TABLE I
PARAMETERS USED

Number of layers	3
Number of hidden layers	1
Number of neurons in hidden layer	30
Number of neurons in output layer	2
Activation function used	sigmoidal
Learning algorithm used	Back propagation and Levenberg-Marquardt

TABLE II
CLASSIFICATION ACCURACY AND RMSE

Techniques	Classification accuracy	RMSE
BPSO -FS - MLP classifier with BP training	97.62	0.1744
BPSO -FS - MLP classifier with LM training	95.24	0.1972
PSO-FS - MLP classifier with BP training	96.43	0.1864
PSO-FS - MLP classifier with LM training	94.64	0.1928

TABLE III
PRECISION AND RECALL FOR FINGER

Techniques	Precision	Recall
BPSO -FS - MLP classifier with BP training	0.974683544	0.974683544
BPSO -FS - MLP classifier with LM training	0.938271605	0.962025316
PSO-FS - MLP classifier with BP training	0.962025316	0.962025316
PSO-FS - MLP classifier with LM training	0.948717949	0.936708861

TABLE IV
PRECISION AND RECALL FOR TONGUE

Techniques	Precision	Recall
BPSO -FS - MLP classifier with BP training	0.97752809	0.97752809
BPSO -FS - MLP classifier with LM training	0.965517241	0.943820225
PSO-FS - MLP classifier with BP training	0.966292135	0.966292135
PSO-FS - MLP classifier with LM training	0.944444444	0.95505618

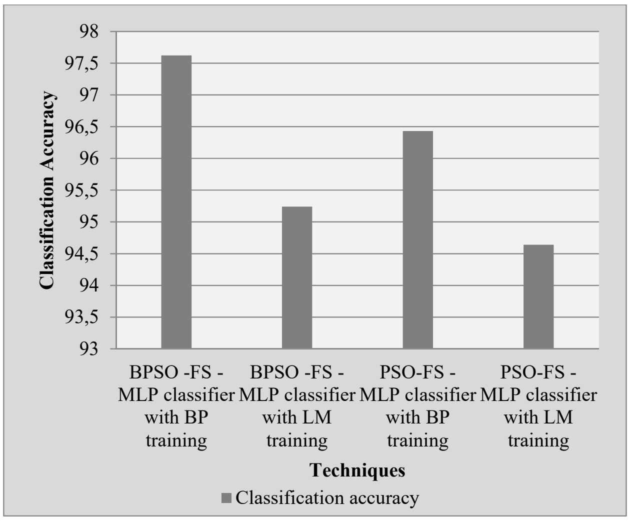

Fig. 3 Classification accuracy

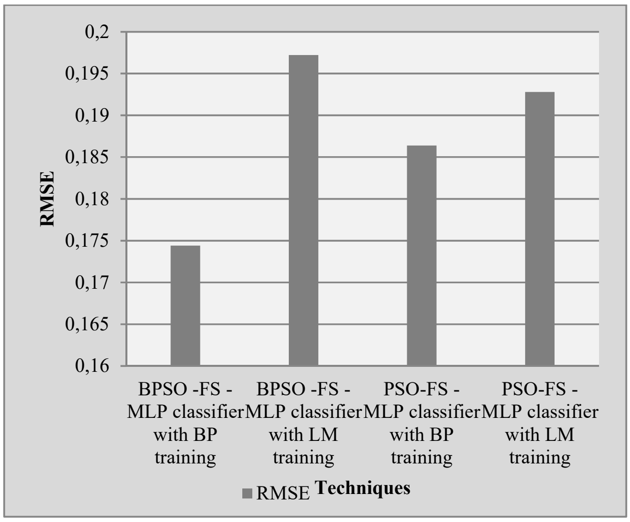

Fig. 4 RMSE

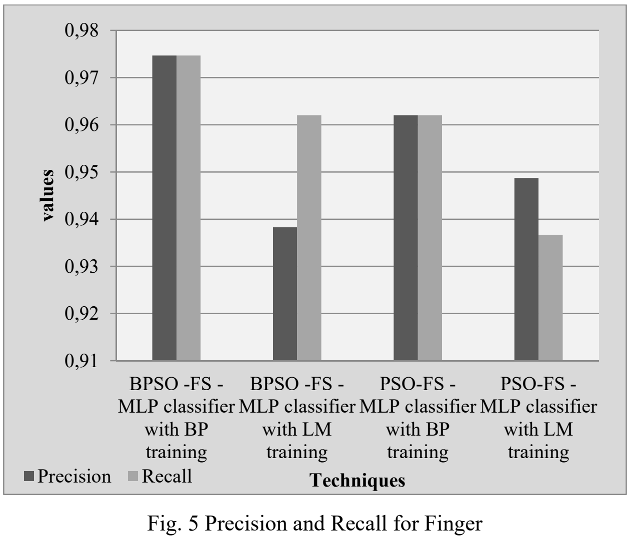

Fig. 5 Precision and Recall for Finger

From Fig. 3, it is observed that the BPSO-FS-MLP with BP algorithm improved accuracy by 1.23% when compared to PSO-FS-MLP with BP algorithm.

From Fig. 4, it is observed that the BPSO-FS-MLP with LM algorithm reduced RMSE by 2.26% when compared to PSO-FS-MLP with LM algorithm.

It can be observed from Fig. 5 that the BPSO-FS-MLP with LM algorithm increased precision by 1.11% than PSO-FS-MLP with LM algorithm and the recall 2.67% for Finger.

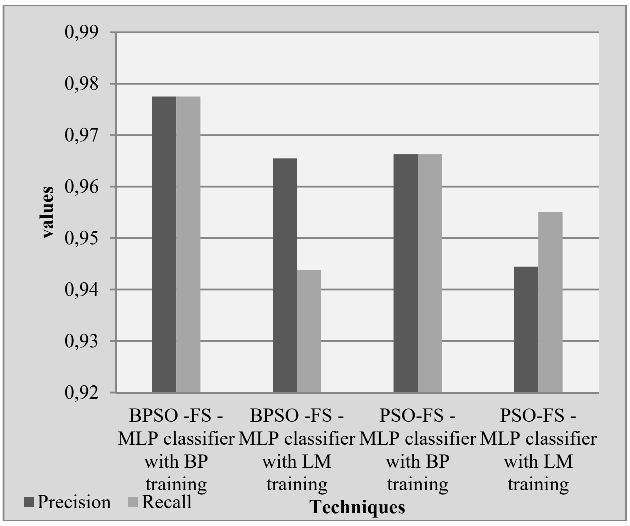

Fig. 6 Precision and Recall for Tongue

It can be observed from Fig. 6 that the BPSO-FS-MLP with BP algorithm increased precision and recall by 1.16% than PSO-FS-MLP with BP algorithm for tongue.

V. Conclusion

BCI's generation performance depends on signal to noise ratio and translation algorithms. Present BCIs have low information transfer rates. Feature selection in pattern recognition involves deriving a feature subset from raw input data to reduce data for classification. This study presented a new feature selection method based on Binary PSO. The proposed algorithm's performance was evaluated with MLP classifier with Back propagation training and Levenberg-Marquardt training algorithms. Results proved that the new method with Back propagation outperformed Levenberg-Marquardt training.

References

[1] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. Clinical neurophysiology, 113(6), 767-791.

[2] Vallabhaneni, A., Wang, T., & He, B. (2005). Brain—computer interface. In Neural engineering (pp. 85-121). Springer US.

[3] Millán, J. D. R., Franzé, M., Mouriño, J., Cincotti, F., & Babiloni, F. (2002). Relevant EEG features for the classification of spontaneous motor-related tasks. Biological cybernetics, 86(2), 89-95.

[4] Khushaba, R. N., Al-Ani, A., & Al-Jumaily, A. (2008, December). Differential evolution based feature subset selection. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.

[5] Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., & Liu, H. (2010). Advancing feature selection research. ASU feature selection repository.
[6] Koprinska, I. (2010). Feature selection for brain-computer interfaces. In New frontiers in applied data mining (pp. 106-117). Springer Berlin Heidelberg.
[7] Schalk, G., & Leuthardt, E. C. (2011). Brain-computer interfaces using electrocorticographic signals. Biomedical Engineering, IEEE Reviews in, 4, 140-154.
[8] Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., & Moran, D. W. (2004). A brain–computer interface using electrocorticographic signals in humans. Journal of neural engineering, 1(2), 63.
[9] Shenoy, P., Miller, K. J., Ojemann, J. G., & Rao, R. P. (2008). Generalized features for electrocorticographic BCIs. Biomedical Engineering, IEEE Transactions on, 55(1), 273-280.
[10] Mishra, A. K., Das, M., & Panda, T. C. (2013). Hybrid Swarm Intelligence Technique for CBIR Systems. International Journal of Computer Science Issues (IJCSI), 10(2).
[11] Abraham, A., Guo, H., & Liu, H. (2006). Swarm intelligence: foundations, perspectives and applications (pp. 3-25). Springer Berlin Heidelberg.
[12] Moubayed, A. (2010). Binary-SDMOPSO and its application in channel selection for Brain-Computer Interfaces. In 2010 UK Workshop on Computational Intelligence (UKCI) (pp. 1-6).
[13] Ang, K. K., Yu, J., & Guan, C. (2012, March). Extracting effective features from high density nirs-based BCI for assessing numerical cognition. InAcoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on (pp. 2233-2236). IEEE.
[14] Polat, D., & Cataltepe, Z. (2012, April). Feature selection and classification on brain computer interface (BCI) data. In Signal Processing and Communications Applications Conference (SIU), 2012 20th (pp. 1-4). IEEE.
[15] Hatamikia, S., Nasrabadi, A. M., & Shourie, N. (2014, November). Plausibility assessment of a subject independent mental task-based BCI using electroencephalogram signals. In Biomedical Engineering (ICBME), 2014 21th Iranian Conference on (pp. 150-155). IEEE.
[16] 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.
[17] Nasehi, S., & Pourghassem, H. (2011, May). 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 on (pp. 114-117). IEEE.
[18] Huang, D., Qian, K., Oxenham, S., Fei, D. Y., & Bai, O. (2011, April). Event-related desynchronization/synchronization-based brain-computer interface towards volitional cursor control in a 2D center-out paradigm. In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on (pp. 1-8). IEEE.
[19] Samiee, S., Hajipour, S., & Shamsollahi, M. B. (2010, June). Five-class finger flexion classification using ECoG signals. In Intelligent and Advanced Systems (ICIAS), 2010 International Conference on (pp. 1-4). IEEE.
[20] Elsawy, A. S., Eldawlatly, S., Taher, M., & Aly, G. M. (2013, September). A principal component analysis ensemble classifier for P300 speller applications. In Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on (pp. 444-449). IEEE.
[21] Yang, Y., Chevallier, S., Wiart, J., & Bloch, I. (2012, August). Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (pp. 2744-2747). IEEE.
[22] Gottemukkula, V., & Derakhshani, R. (2011, April). Classification-guided feature selection for NIRS-based BCI. In Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on (pp. 72-75). IEEE.
[23] Gonzalez, A., Nambu, I., Hokari, H., Iwahashi, M., & Wada, Y. (2013, October). Towards the Classification of Single-Trial Event-Related Potentials Using Adapted Wavelets and Particle Swarm Optimization. In Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on (pp. 3089-3094). IEEE.
[24] Bhattacharyya, S., Rakshiti, P., Konar, A., Tibarewala, D. N., Das, S., & Nagar, A. K. (2013, April). Differential evolution with temporal difference Q-learning based feature selection for motor imagery EEG data. In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2013 IEEE Symposium on (pp. 138-145). IEEE.
[25] Shahriari, Y., & Erfanian, A. (2011, April). A mutual information based channel selection scheme for P300-based brain computer interface. In Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on (pp. 434-437). IEEE.
[26] Jayathilake, A. A. C. A., Perera, A. A. I., & Chamikara, M. A. P. Discrete Walsh-Hadamard Transform in Signal Processing.
[27] Johnson, J., & Puschel, M. (2000). In search of the optimal Walsh-Hadamard transform. In Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on (Vol. 6, pp. 3347-3350). IEEE.
[28] Sasikala, D., & Neelaveni, R. (2010). Correlation coefficient measure of multimodal brain image registration using fast walsh hadamard transform. Journal of Theoretical & Applied Information Technology, 22(2).
[29] Abdolreza Asadi Ghanbari, Karim Adinehvand, & Mousa Mohammad Nia (2014) Overhead Reduction in EEG signals using Particle Swarm Optimization and Independent Component Analysis
[30] Hasan, B. A. S., & Gan, J. Q. (2009). Multi-objective particle swarm optimization for channel selection in brain-computer interfaces.
[31] Das, S., Abraham, A., & Konar, A. (2008). Particle swarm optimization and differential evolution algorithms: technical analysis, applications and hybridization perspectives. In Advances of Computational Intelligence in Industrial Systems (pp. 1-38). Springer Berlin Heidelberg.
[32] Balochian, S., Seidabad, E. A., & Rad, S. Z. (2013). Neural Network Optimization by Genetic Algorithms for the Audio Classification to Speech and Music. International Journal of Signal Processing, Image Processing & Pattern Recognition, 6(3).
[33] Taravat, A., Proud, S., Peronaci, S., Del Frate, F., & Oppelt, N. (2015). Multilayer Perceptron Neural Networks Model for Meteosat Second Generation SEVIRI Daytime Cloud Masking. Remote Sensing, 7(2), 1529-1539.
[34] Lotte, F., Congedo, M., Lécuyer, A., & Lamarche, F. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. Journal of neural engineering, 4.
[35] Craven, M. P. (1997). A faster learning neural network classifier using selective backpropagation.
[36] CH Satyananda Reddy (2012) design of multilayer perceptron neural network for mental task recognition, 1(1), 74-80.
[37] Yu, H., & Wilamowski, B. M. (2011). Levenberg-marquardt training. Industrial Electronics Handbook, 5, 12-1.
[38] Sapna, S., Tamilarasi, A., & Kumar, M. P. (2012). Backpropagation learning algorithm based on Levenberg Marquardt Algorithm. DOI= DOI, 10.
[39] Mojtaba Ahmadieh Khanesar, Hassan Tavakoli, Mohammad Teshnehlab and Mahdi Aliyari Shoorehdeli (2009). Novel Binary Particle Swarm Optimization, Particle Swarm Optimization, Aleksandar Lazinica (Ed.),
[40] Khanesar, M. A., Teshnehlab, M., & Shoorehdeli, M. A. (2007, June). A novel binary particle swarm optimization. In Control & Automation, 2007. MED'07. Mediterranean Conference on (pp. 1-6). IEEE.

K. Akilandeswari is Research Scholar in Periyar University, Salem. She is currently pursuing her doctorate in India.

G. M. Nasira is with Chikanna Government Arts College, Tirupur, India.