Classification of Mental Tasks using EEG and Hierarchical Classifier employing Optimised Neural Networks

Madhuri N. Bawane
Dept. of Electronics & Telecomm. Engg.
Government Polytechnic
Nagpur, India

K. M. Bhurchandi
Dept. of Electronics and Communication Engg.
Visvesaraya National Institute of Technology
Nagpur, India

ABSTRACT
With recent advances in Electroencephalogram (EEG) signal processing and biomedical instrumentation, brain machine interfaces are used for rehabilitation of people suffering from neuromuscular disorders. This paper presents a novel method employing Hierarchical classifier using optimised Neural Networks to classify left-hand movement, right-hand movement and word generation using EEG signals. One of the most important components of brain computer interface (BCI) is feature extraction of EEG signals. Power spectral density (PSD) is used for feature extraction from EEG signals. The proposed pre-processing and reconfiguration of PSD samples make them more discriminative & yield appropriately organized feature vectors. The adaptation of network weights using Comprehensive Learning Particle Swarm Optimization (CLPSO) is proposed to improve the performance of Neural Network (NN). Further, the two level hierarchical neural network is used to enhance the discriminative property of the features and hence better classification accuracy is achieved. Results are verified on BCI benchmarking database as well as our own experimental database. Results obtained using the proposed methods are compared with other contemporary methods such as Linear Discrimination analysis (LDA), neural networks based on improved particle swarm optimization (IPSONN) and to a recently proposed approach based on Evidence-based combining classifier. It is found that the proposed method outperforms all the contemporary techniques for the multi-task EEG classification. This new method can be easily extended to other multitask BCI applications.

General Terms
BCI, EEG Classification, Neural networks

Keywords
Hierarchical Classifier, Successive input resampling, CLPSO

1. INTRODUCTION
Advances in cognitive neuroscience and brain imaging technology have provided us an ability to interface directly with the human brain activities. The physical processes in the brain that correspond with certain forms of thoughts can be monitored to a limited level through these technologies. Researchers have used this fact to build brain computer interfaces (BCI) i.e. communication systems that do not depend on normal pathways of peripheral nerves and muscles [1].

The effectiveness of BCI is investigated in the field of rehabilitation [2] and neurobotics [3-5]. BCIs measure activities of brain, process them and produce information and control signals that reflect the user's intentions.

Electroencephalogram (EEG) signals measured from the scalp serve as an easy non-invasive method for measuring brain activities [6]. Numerous studies have demonstrated correlation between EEG signals and actual or imagined movements. EEG measures the electrical potential with the help of several electrodes at predefined points on a human skull with one additional electrode (sometimes between the eyes above the nose, sometimes behind one of the ears) to ground the subject. Moving a limb or even contracting a single muscle affects brain activity in the cortex.

Imagination of movements is called sensory motor rhythm (SMR). Sensory Motor Rhythms are the oscillations in brain activity signals recorded from somatosensory and motor areas [7]. Brain oscillations are categorized according to specific frequency bands which are named after Greek letters (delta: < 4 Hz, theta: 4–7 Hz, alpha: 8–12 Hz, beta: 12–30 Hz and gamma: > 30 Hz). Alpha band is also referred as mu band. The decrease of oscillatory activity in a specific frequency band is called event-related desynchronization (ERD) and the increase of oscillatory activity in a specific frequency band is called event-related synchronization (ERS). ERD/ERS patterns are produced by motor imagery which is the imagination of movements without actually performing the movement. The frequency bands mu and beta in EEG signals are significant for motor imagery.

Conventional and translation-invariant (TI) wavelet-based approaches for single-trial evoked potential estimation based on intra-cortical recordings are presented in [8]. The excellent performance of the wavelet-based approaches for extracting evoked potentials is also demonstrated in this paper using examples on simulated and experimental data. The method proposed in [9] selects the coefficients based on the statistical study of trials from the training data set and uses wavelet coefficients of the signal for feature detection. In [10], a radial basis function neural network (RBFNN) is used to extract event-related potential (ERP) using the EEG signal. This paper claims that RBFNN is useful for real-time detection of attention-related ERP and thus can be an effective tool for the ERP amplitude and latency change estimation. Various feature extraction methods have been applied to obtain features from the EEG signals including adaptive autoregressive models (AAR) [11-13], common spatial filters[14,15] and wavelet transform [16 , 17].

The t-CWT is a novel method for feature extraction from biological signals and is introduced in [18]. It is based on continuous wavelet transform (CWT) and Student's t-statistic applied to event-related potential data in brain computer interface. This method provides fully automated detection and quantification of ERP components that best discriminate between two samples of EEG signals. A new method is also

proposed in [19] to reduce the number of EEG channels needed to classify mental tasks. Genetic algorithm is used for the search space consisting of 6 channel combinations of 19 EEG channels. More salient combinations of them in classification of three mental tasks are selected. Different classification methods have been proposed by various researchers to analyze EEG signals. A prominent few of them can be enlisted as: Linear discrimination analysis (LDA) [20], Self organizing Feature Map [21], Multiple Kernel Learning Support Vector Machine [22], neural networks with improved particle swarm optimization [23], recurrent neural networks [24], time-frequency analysis [25], Fuzzy sets based classification [26] and Independent Component Analysis (ICA) [27].

This paper presents a novel but a very simple method to effectively classify the EEG of mental tasks for left-hand movement imagination, right-hand movement imagination and word generation. The three mental tasks classified correspond to three control commands in a BCI system. Pre-computed power spectrum density (PSD) features detected using Welch's periodogram [28] are used. The PSD values are then pre-processed before applying to the proposed classifier. Further, we classify these features using variants of optimised back propagation Neural Network on the proposed Hierarchical classifier. The results are compared with other published contemporary techniques.

The outline of this study is as follows. Section 2 describes PSD computation, particle swarm optimization and basic classifiers required for the proposed work. The proposed Data pre-processing, optimization using CLPSO and hierarchical classification using Neural Network is described in section 3. In section 4, experimental methodology using the BCI dataset and B-alert experimental setup used for the experiments on the EEG signals are presented. Section 5 presents and benchmarks the achieved performance. Finally the study is concluded in section 6.

2. METHODS AND MATERIALS

This section presents in brief the fundamental of the theory and principles required for the proposed work.

2.1 Power Spectral Density

Power Spectral Density (PSD) is a positive real function which describes the power distribution of a signal over frequency. It represents the average power distribution in frequency domain i.e. as a function of frequency. Several parametric and non-parametric approaches for estimation of this distribution are available. In a common parametric technique, an autoregressive model is fitted to the observations. Welch's periodogram is an effective non-parametric technique. It is used in the calculation of PSD features that are further classified using various classifiers. Welch's procedure to estimate the PSD of a signal combines windowing and averaging in order to obtain a smooth spectrum without random fluctuations resulting from the acquisition and estimation processes. K possible overlapping segments are formed by dividing original data sequence of each channel. A non-overlapping window v[n] is defined over each of these segments and the corresponding periodograms are computed and averaged. If $x^{(k)}[n]$ represents the sample x[n] of the kth data segment of length N, then the periodogram for that segment is defined as

$$\hat{P}_x^{(k)}(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} v[n] x^{(k)}[n] e^{-j\omega n} \right|, \quad k = 1, ..., K \quad (1)$$

where $\omega = 2\pi f$ (in rad/s) is the angular frequency and the window v should obey the normalization

$$(1/N) \sum_{n=0}^{N-1} v[n]^2 = 1 \quad (2)$$

Then the PSD of the signal, for each frequency ω is given as,

$$\hat{S}_x(\omega) = \frac{1}{K} \sum_{k=1}^{K} \hat{P}_x^{(k)}(\omega) \quad (3)$$

Using precomputed PSD as a feature vector either generated from the BCI data set or acquired from the B-Alert system, we further propose overlapped windowing of the PSD signals.

2.2 Particle Swarm Optimization

Particle swarm optimization (PSO) is a robust nature inspired algorithm for solving optimization problems based on movement and intelligence of swarms [29]. It searches for the best solution by simulating the movement and flocking of birds. The algorithm works by initializing a flock of birds randomly over the searching space where every bird is called as a ''particle''. Each particle is treated as point in D dimensional space. These ''particles'' move with an adaptable velocity and adjusts its own position according to its own experience and that of neighbors. Each particle keeps track of its position in search space that is associated with the best solution achieved by it. This is called personal best; pbest. The swarm remembers another value denoted, *gbest*, is the best position discovered so far by the swarm. The trajectory of each particle in the search space is dynamically adjusted by updating its velocity, according to its pbest and gbest. Therefore, PSO combines the local search technique and the global search method to move towards the global optimum. The closeness of particle to this optimum is measured according to fitness function for the problem to be solved. In this work, particle represents weight vector of neural networks including biases. The dimension of search space is total number of weights and biases.

As the particle moves in the search space, the position of ith particle is represented as $X_i = (x_{i,}^1, x_{i,}^2, x_{i,}^3, \cdots, x_{i,}^D)$ and pbest of ith particle is pbest $= (p_{i,}^1, p_{i,}^2, p_{i,}^3, \cdots, p_{i,}^D)$. The velocity is described by $V_i = (v_{i,}^1, v_{i,}^2, v_{i,}^3, \cdots, v_{i,}^D)$. Each particle is attracted to the best position encountered by it so far and overall best position found by neighborhood with increased iteration. The global best gbest is represented as gbest $= (g_i^1, g_i^2, g_{i,}^3, \cdots, g_i^D)$.

The velocity and position update equation of dth dimension is given by

$$V_i^d(t+1) = w V_i^d(t) + c_1 * rand1^d() * (pbest_i^d(t) - X_i^d(t)) + c_2 * rand2^d() * (gbest_i^d(t) - X_i^d(t)) \quad (4)$$

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1) \quad (5)$$

where t indicates a pseudo time (iteration increment), w is inertia weight, c_1 and c_2 are acceleration constants. $rand1^d$ and $rand2^d$ are random numbers in the range of [0 1].

2.3 Classifiers

Performance of various Feed forward back propagation networks with various training algorithm are examined for classification accuracy. An example of four layer Multi layer
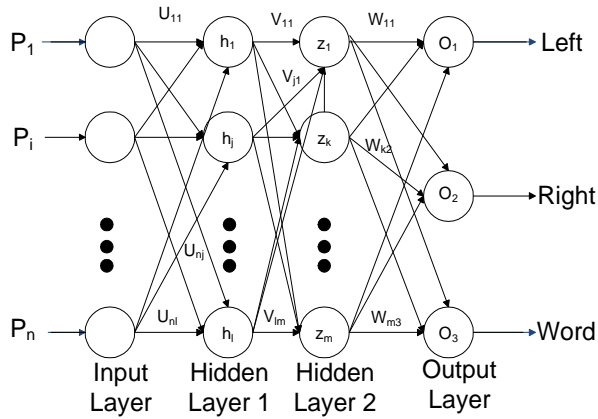
Figure 1. Feed Forward ANN with 02 hidden layers

Perceptron (MLP) with two hidden layer and one output layer is shown in Figure 1. As already said, we have used a simple MLP to achieve the demonstrated results as against the most complex classifiers used in [20-27]. Feed forward neural network like Multi layer Perceptron (MLP) with different variants of back propagation learning algorithms assisted by CLPSO are used to achieve better classification. To overcome limitations of basic MLP in multi-task classification and to further improve result, novel concept of two level Hierarchical classifier is proposed which uses multiple MLPs with simple architecture to achieve nonlinear decision boundaries that result in more robust classification. Though MLPs are the most widely used neural networks in classification of BCI signals, the modification into hierarchical classifier offers them the robustness required for accurate and reliable classification required to solve problems such as mental multi-task classification.

3. PROPOSED WORK
In this section, a typical way of preprocessing and applying the acquired samples for analysis of different mental activities named as 'Successive input resampling' and a novel method for neural network classification of EEG signals named 'Hierarchical classifier' is proposed. Optimization of the neural network is achieved using CLPSO.

3.1 Successive Input Resampling
Our actions are planned and initiated by our mental processes (many times unconscious) before one actually initiate execution of tasks. When a person is going to perform a movement, the body runs a chain of events that ends with the action of the muscles and therefore, the actual movement [30]. This sequence starts in the brain only a few milliseconds before the movement inception, and subsequently, the electrical signal passes through spinal cord and reaches the muscles that exert the necessary force. Thus planning of the next activity and execution of the current activity is overlapped. To implement this, overlapping of the successive windows while averaging is experimented. Averaging of 16 samples with 50% overlap i.e. 8 samples overlapping is applied as shown in Figure 2. This strengthens the correlation between the two successive samples of 16 discrete readings.
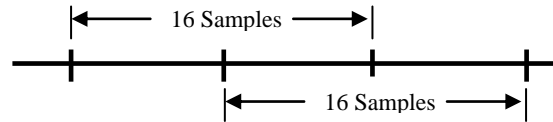

Figure 2. 50% Overlapping of samples

16 sample window with 50 % overlap yields almost equal samples as for the 2 Hz data. This new data set is applied to the hierarchical classifier proposed in the next section.

3.2 Hierarchical Classifier
The hierarchical classification is a technique for improving the classification accuracy of a MLP classifier system. An unclassified data is considered at the root (down) of the tree. It involves application of more focused and targeted classifiers as we traverse up the classification tree. Individual stages up the tree perform more and more difficult classification tasks. Hence, in the first level at the root, the input features are classified into broad categories may be resulting in classification of one or more clearly classifiable input and also one or more groups of misclassified inputs. The correctly classified inputs are retained and the second level will have multiple autonomous classifiers to classify each misclassifying or confusing class into possible correctly classified inputs. Successive stages will use the misclassified groups at each output level and yields appropriate classes as outputs of the next level as in Figure 3. This approach results in a correct but coarse classification initially and makes it finer successively, finally resulting in robust classification.

Let L represent set of classes C_k to be discriminated. If we divide L into M non-empty subsets L_i so that members of L_i are least confused with members of L_j ($\forall_j \neq i$). A specific class C_k will be a member of L and one of the subsets L_i. Hence posterior probability of class C_k as a joint probability of the class and the respective subset L_i can be written as in (6).

$$p(C_k|x) = p(C_k, L_i|x), C_k \in L_i$$
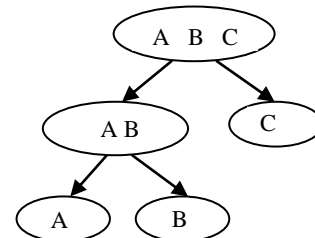$$= p(L_i|x)\, p(C_k|L_i, x) \quad (6)$$

Figure 3. A Hierarchical Classifier

The task of discriminating between all the classes in L can be transformed into discriminating subsets L_i and separately distinguishing the classes C_k remaining in each of the subsets L_i. Repetition of this procedure results in a hierarchical tree like classifier structure. The training data for each node is shared among all its child nodes as per the separation of classes. The training data reduces up the tree with increase in accuracy of the classification.

It is found that for 3 class BCI system, the accuracy is limited due to close resemblance of the feature vector for right hand

movement and word generation task. In the proposed work, two level hierarchical classifier using neural network as shown in Figure 4 is used to enhance the discriminative property of the features and hence achieve robust classification and better classification accuracy. The input feature vector is applied to ANN1 that discriminates left hand movement from right hand movement and word generation features. We used three layer feed forward artificial NN with one hidden layer containing 30 neurons. The target values are set at 0 (Left hand movement) and 1 (Right hand movement and word generation). The output of ANN1 is used to activate ANN2. Again the original input feature vector is fed to second
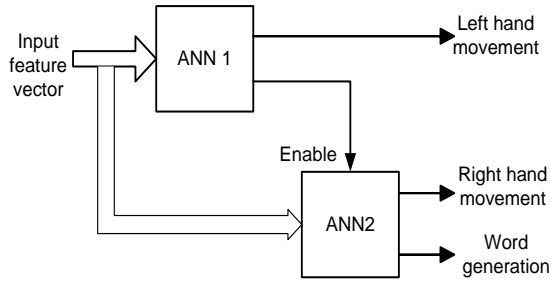

Figure 4. Hierarchical Classifier Using Neural Network

neural network ANN2 that is architecturally similar to ANN1. ANN2 now discriminates right hand movement and word generation task. This network also uses single hidden layer with 30 neurons. CLPSO is used to find initial weights and biases for neural networks used in   proposed Hierarchical classifier system and thus to optimize performance of neural network under consideration. Performance of proposed classifier is cross-validated on the BCI competition and B-Alert datasets with the PSD features.

3.3 Comprehensive Learning Particle Swarm Optimization

A comprehensive learning strategy is employed to improve the PSO's capability to deal with complex problems and specifically aimed at overcoming the problem of premature convergence [31]. In this algorithm, each particle's velocity vector can be updated by using not only its own best but also any other particles pbest. The particles velocity is updated according

$$V_i^d(t+1) = \mathrm{w}V_i^d(t) + c_1 * rand1^d() * (pbest_{f_i(d)}^d(t) - X_i^d(t)) \quad (7)$$

where $f_i = [\ f_i(1),\ f_i(2),\ \text{------}\ f_i(D)]$ decides the pbest that the particle i should follow. In the learning strategy, each particle learns from all particle's pbests in the swarm. During the search process, each dimension of a particle has an equal chance to learn from other particles. The decision depends on probability Pc that takes different values for different particles. The expression for Pc is given by [31]

$$Pc_i = 0.05 + 0.45 * \frac{\exp\left(\frac{10(i-1)}{s-1}\right) - 1}{\exp(10) - 1} \quad (8)$$

where i is particle's id and s is population size.

The algorithm for CLPSO is as follows:

   (1) Initialize the swarm of p particles such that positions and velocities are random in the range.

   (2) Evaluate learning probability Pc_i for all particles using equation (8).

   (3) Evaluate the performance FXi for each particle and update its pbest . Also find gbest.

   (4) For each particle do

     4.1 For d=1 to D

  (4.1.1) Generate a random number in the range [0 1] If this number is greater than Pc_i, $f_i(d) = i$

  (4.1.2) Else randomly choose two other particles and select the one with better fitness value.

  (4.1.3)Update the particle velocity according to Equation (7)

 (4.2) Move each particle to new position according to Equation given below

$$X_i^d(t+1) = X_i^d(t) + rt * V_i^d(t+1) \quad (9)$$

where rt is velocity retardation factor [32]

   (5)   Calculate w using equation

$$w = w_{max} - \left(\frac{(w_{max} - w_{min}) * iteration\_no)}{Max\_iteration}\right) \quad (10)$$

  (6) Continue till Max iterations or mean square error (MSE) condition is satisfied

  (7) Go to step 3

4. EXPERIMENTS

This section presents the discussions on the online BCI benchmark datasets and our own acquired data using B-Alert system and the subsequent experimentation. It also describes the signal pre-processing, experimental methodology and the experiments themselves before moving on to the results and discussion section.

4.1 EEG Datasets

In this work, Dataset V of BCI competition III is used. This dataset is provided by IDIAP Research Institute presented by Silvia Chiappa and et al. [33]. The dataset has data from three normal subjects recorded during four non-feedback sessions. The duration of each session was four minutes with break of 5-10 minutes. All sessions of a particular subject were recorded on the same day. The subject executed a specified task for about 15s and then switched randomly to another task as per the instructions. The continuous sample readings were taken as depicted in Figure 5 in the intervals of 15 seconds each. Three mental tasks were classified in this study using the acquired data.

   1.   Imagination of the left-hand movements (left).

   2.   2. Imagination of the right-hand movements (right).

   3.   Generation of words beginning with the same random letter   (word).

A cap with 32 integrated electrodes positioned at standard locations of the International 10–20 system was used to record the EEG signals. The sampling rate was 512 Hz.

Experiments were performed to create a second dataset using B-Alert system. B-Alert system is a wireless EEG data acquisition system. The data has been recorded using B-Alert system. The subjects were given to perform three different tasks on the same lines as discussed above for the BCI competition data set so that the designed hierarchical classifier should be seamlessly able to switch between the two datasets.

4.2 Methodology and Performance Evaluation

The experiments were carried out for classifying three different tasks. All the data were obtained from the BCI experiment described in Section 4.1. The raw EEG data was preprocessed as discussed in section 3.1. The result was obtained by using the preprocessed PSD and the proposed hierarchical classifier. For a given subject, four continuous non-feedback sessions were recorded. A surface Laplacian filter was used to filter the EEG signals[33]. The power spectral density (PSD) in the 8–30 Hz band was calculated every 62.5ms with a frequency resolution of 2 Hz. The eight centro-parietal electrodes used were C3, Cz, C4, CP1, CP2, P3, Pz, and P4. Each session contains roughly equal number of samples for left, right hand movement and word generation. The exact locations of the eight channel sensors used are shown in Figure 6. Welch periodogram method was used for estimation of power spectral density [34]. Thus EEG signals are represented in the form of 96 dimensional feature vectors with every 12 entries coming from each of eight electrodes sequentially.
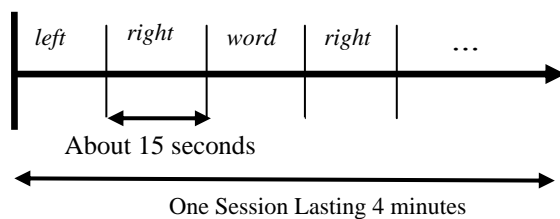

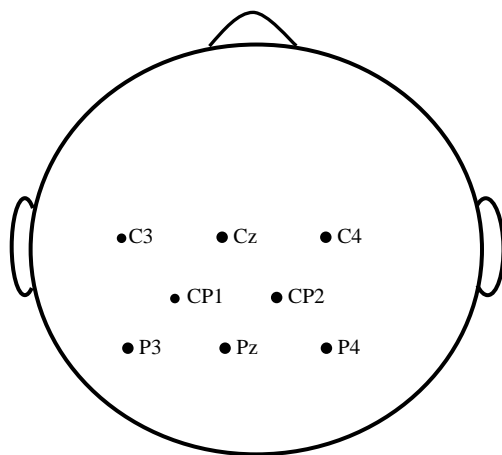**Figure 5. Data Acquisition Pattern for One session**

Figure 6. Electrode positions for data collection

4.2.1 Methodology with BCI Dataset

Various types of classifiers such as MLP, MLP with CLPSO, simple Hierarchical and Hierarchical with CLPSO are used and classification accuracy is computed using successive trials. The sets of training and testing samples used for performance evaluation were mutually exclusive. Each training session was followed by multiple testing trials. Each testing trial consisted of typically 100 classification efforts. The classification efficiency was computed as a ratio of the number of correctly classified samples in each trial to the number of total classification attempts in the trial. Several such classification trials were carried out and average of all the individual trials were computed as final percentage

classification accuracy. The following patterns of preprocessing were used for the classification.

Method 1- Input feature vector (12 temporal samples x 8 sensors=96 features) acquired 16 times per second (16 Hz data) as available in BCI dataset is considered for training. It provides acquired feature vectors for three sessions of approximately equal duration. The dataset of the three sessions consists total 10528 vectors of 96 features each with the respective labels. Each feature vector corresponds to one of the three mental actions.

Method 2 and 4 - The average of non overlapping 8 consecutive samples of 16 Hz data is calculated in order to produce 2 Hz data. This yields number of training samples 10528/8= 1316 for all the three sessions. The method 2 uses non overlapped window averaging using conventional MLP classifier while the method 4 uses hierarchical classifier.

Method 3 and 5 – EEG Signals for successive mental activities overlap practically. Overlapping windows of successive 16 samples are taken for computing average all over the data. The successive windows overlap of 50 % is selected for this experimentation. This results in 1313 training samples for all the sessions taken together. The method 3 uses overlapped window averaging using conventional MLP classifier while the method 5 uses hierarchical classifier.

4.2.2 Classifier Training

A MLP NN with two hidden layers and one output layer is used for the conventional single MLP classifier to separate the three classes. It is trained using various training algorithms to classify three mental tasks with processed PSD as features. A network is selected with 30 neurons in the two hidden layers and 03 Neurons in the output layer and the target values were set to 1,0,0 for the Left hand movement imagery, 0,1,0 for Right hand movement imagery and 0,0,1 for Word Generation.

Number of test samples for 16 Hz given in the BCI competition are 3504. The test samples for 2 Hz data are 438 while for 50 % overlap the test samples are 435 for the individual training of three sessions. Combinations of tan sigmoid and Log sigmoid activation functions were experimented in all the layers with 3 neurons in output layer and starting from 3 neurons in hidden layer. The number of neurons in the first hidden layer was experimented up to 45. It did not improve the classification accuracy. Performance of the network is verified for variety of neurons in hidden layers. During the experimentation it is observed that the network with 30 neurons in each hidden layer yield the best classification accuracy for various training algorithms. Similarly tan sigmoid functions in all the three layers of the neural networks provided the best percentage classification accuracy as compared to other activation functions.

The parameters set for the CLPSO [31] to calculate initial weight and bias of network are as follows:
Population size s = 30
Maximum Iterations = 30
Acceleration constant c_1= 1.49445
Initial Inertia weight w = 2
Maximum inertia weight w_{max} = 0.9
Minimum inertia weight w_{min} = 0.5
Velocity retardation factor rt = 0.8

The weights and biases arrays are initialized at random to be the particle position in the search space. The iterations are

fixed and particles are displaced in space for fixed number of iterations. The weights and biases at the last iteration corresponding to the personal best are then taken as initial weights for the neural network.

4.3 Methodology with B-alert Setup

For dataset on B-Alert system, data was recorded for all tasks with sampling frequency of 256 Hz and then resampled to 16Hz data. Data was recorded on the same eight channels as BCI dataset. The three tasks recorded again are Imagination of left hand movement, imagination of right hand movement and generation of words beginning with same random letter.

The B-Alert Software (BAS) acquires, presents, and stores physiological signals from the device. B-Alert machine also has functionality for retransmitting data, computing and displaying it in real-time cognitive metrics, administering baseline cognitive assessments and replaying data offline. Working setup of wireless BCI system is shown in Figure 7. The data outputs are saved in the universally compatible EDF+ format. PSD is estimated using Welch method in the frequency band 8-30 Hz with a frequency resolution of 2 Hz.
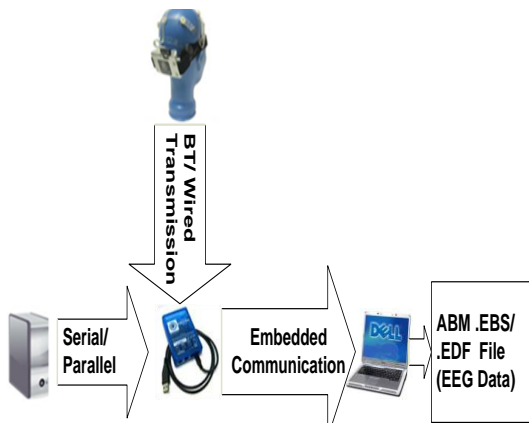

Figure 7(a). B-Alert Experimental Set up

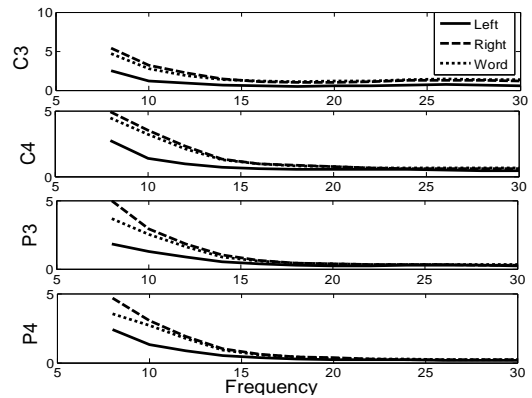**Figure 7(b) EEG Recording with B-Alert System**

Figure 8. PSD for the Major Selected Channels having significant signal variations

The universally compatible EDF+ format. PSD is estimated using Welch method in the frequency band 8-30 Hz with a

This gives the 12 PSD components for each channel providing 96 PSD components obtained for 8 channels. The PSD graph for selected channels is plotted in Fig. 8.

5. RESULTS AND DISCUSSION

Novel concept of Hierarchical classifier is introduced in this paper to overcome limitations of MLP and existing methods while dealing with multi-task classification problem. To enhance performance of this classifier, CLPSO based neural networks are used as basic element in this classifier.

The total samples for training for all the three sessions were 10528 as available in the BCI dataset. The test samples for 16 Hz data as available in the dataset are 3504. The achieved accuracy for this 16 Hz data was 71.77 %. Then the average of 8 consecutive samples is calculated in order to produce a response at every 0.5 seconds (2 Hz data) yielding 1316 samples for training. The average of non overlapping 8 consecutive samples of 16 Hz test data is calculated in order to produce 2 Hz data. This yields number of test samples 3504/8 = 438. This yielded accuracy of 77.33% for non overlapping 2 Hz data. The overlapping of brain and muscular activities for the successive physical actions and the proposed overlapping of subsequent EEG signal patterns is the newly proposed idea and original contribution in this work. This is termed as 'Successive input Resampling ' that is introduced for pre-processing of features in this paper. Overlapping windows of 16 consecutive samples with overlap of 50% i.e. eight samples are used and average of each window is computed resulting in total 1313 training samples for the three sessions and 435 test samples. With MLP, accuracy obtained using overlapped samples for training is 79.64 %.

It was observed that the accuracy of left hand movement was better than the other two tasks. As per the study, the right hand movements and word generation task are both controlled by the same halve of brain. Hence the classifier is explored that will first separate out the tasks controlled by left and right hemisphere of brain. A hierarchical classifier is used where individual stages perform more difficult classification as one go down in the hierarchy. Thus in the first level input features are classified into broad categories. The second level of the hierarchical classifier then segregates the group that contains right hand movement and word generation task.
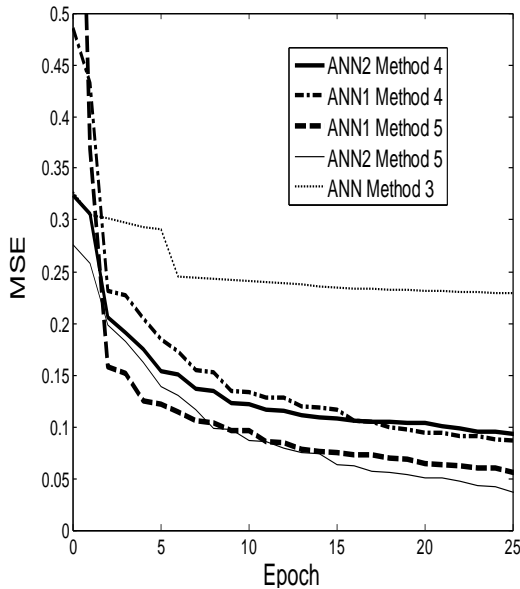
Figure 9. MSE plots for method 3 (overlapping data, single MLP), method 4 (non-overlapping data, hierarchical classifier) and method 5 (overlapping data, hierarchical classifier)

The 8 sample averaged (2Hz) data was first applied to the Hierarchical classifier with one hidden layer in both level of neural networks. The best accuracy with this combination of features and classifier was 81.8 %. Then the hierarchical classifier was used with 16 sample window average with 50% overlap and the accuracy obtained was 84.32 %. Neural network training with backpropagation algorithms have very slow convergence and may trap in the local minimum. CLPSO is used for deciding the initial weights and biases for initialization of the MLP neural network. The parameters required for implementation of the CLPSO algorithm are the population size, number of iteration, w, w_{max} w_{min} , rt and c_1. The value of these parameters are specified in section 4.2.2. The classification accuracy for hierarchical classifier with optimized neural network is improved to 86.85 %. The classification results for various methods of data resampling and using the various classifiers discussed earlier are summarized in Table 1. It is obvious that the proposed data resampling and the hierarchical classifier both together achieve considerable improvements in the classification accuracy. The mean square error plots of hierarchical classifier network components ANN1 and ANN2 for different methods have been presented in Figure 9.

The performance comparison of various published methods using the same dataset are presented in Table 2. The proposed Hierarchical Classifier with input resampling is compared with different so far published methods proposed by C. Lin and M. Hsieh that use neural networks based on Improved

Table 1. Classification Accuracy of various methods on BCI dataset

Method	Data Preprocessing	Classifier	Feature Vector size	% Accuracy
1	BCI dataset	Single MLP	96x3504	71.77
2	8 sample	Single	96x438	77.33

	window average	MLP		
3	16 samples window average with 50% overlap	Single MLP	96x435	79.64
4	8 samples window average	Hierarchical	96x438	81.8
5	16 samples window average with 50% overlap	Hierarchical	96x435	84.32
6	16 samples window average with 50% overlap	Single MLP with CLPSO	96x435	84.21
7	16 samples window average with 50% overlap	Hierarchical classifier with CLPSO	96x435	86.85

particle swarm optimization [23], Support vector machine (SVM)[35], Linear discriminant analysis (LDA) [35] to the recently used Evolved Filters with SVM [36] and an evidence-based combining classifier [37] in Table 2. It is clear that, the proposed methods outperform all the published methods. Results of all the proposed methods on the dataset collected in our laboratory using the B-Alert system with the similar protocols are presented in Table 3. Table 4 presents comparison of different classifiers for B-Alert dataset. It is clear that the proposed classifier is equally robust and effective on our own dataset for classification of the same tasks.

Table 2: Comparison between different classifiers for BCI data

S.N.	Classification Method and Reference	Classification Accuracy (%)
1	IPSONN [23]	78.31
2	SVM [35]	78.08
3	LDA[35]	77.40
4	Evolved Filters (SVM) [36]	79.97
5	Evidence-based combining classifier[37]	83.1
6	**Proposed Hierarchical Classifier**	84.32
7	**Proposed Hierarchical Classifier with CLPSO NN**	**86.85**

Table 3: The classification accuracy using various methods on our B-Alert Experimental Data

Feature processing	Averaging without overlap	50 % overlap	Averaging without overlap	50 % overlap
Classifier	Single MLP		Hierarchical	
% Average Accuracy	81.33	83.25	85.37	**87.52**

Table 4: Comparison between different classifiers for B-Alert data

S.N.	Classification Method	Classification Accuracy (%)
1	SVM	84.81
2	LDA	77.25
3	BPNN	83.25
4	**Proposed Hierarchical Classifier**	87.52
5	**Proposed Hierarchical Classifier with CLPSO NN**	90.55

6. CONCLUSIONS

The published state of the art classifier methods offer limited performance for classification of the EEG signals in spite of complex computational requirements. The proposed method using Overlapped Input Resampling and Hierarchical Neural Network classifier offers significant improvement in the EEG classification in terms of classification accuracy. It is observed from various experimental results that the resampled input feature PSD vectors and the hierarchical classifier both individually contribute to the improvement. Though Performance of the state of the art methods degrade in case of a multi-class EEG classification problem, the proposed technique offers a robust classification accuracy even in case of multi-class classification.

The hierarchical classifier outperforms a single MLP classifier. Because of the improved accuracy over existing methods using even less number of hidden layers as compared to MLP NN, the proposed method is found very promising for multi-class classification for BCI system. The performance of hierarchical classifier is improved using modified CLPSO based technique to properly initialize weights and biases of the neural networks.

Future works may consider improving the classification accuracy using more sophisticated classifiers optimized by multi objective techniques for optimal network architecture. The number of hidden layers, neurons and novel techniques of organizing the EEG data to achieve improved performance may be explored.

7. REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, T.M. Vaughan,2000. Brain–computer interface technology: A review of the first international meeting, IEEE Trans. Rehab. Eng. 8, 164–173 2000.

[2] Wolpaw J.R., McFarland D.J., Vaughan T. et al, "The Wadsworth Center brain–computer interface (BCI) research and development program," IEEE Trans Neural Syst Rehabil Eng, Vol. 11, pp. 204-207, 2003.

[3] Kositsky M., Karniel A., Alford S. et al, "Dynamical dimension of a hybrid neurorobotic system," IEEE Trans Neural Syst Rehabil Eng, Vol. 11, pp. 155-159, 2003.

[4] Moore M., "Real-world applications for brain–computer interface technology," IEEE Trans Neural Syst Rehabil Eng, Vol. 11, pp. 162-165, 2003.

[5] Taylor D., Tillery S., Schwartz A., "Information conveyed through brain-control: cursor vs robot," IEEE Trans Neural Syst Rehabil Engg, Vol. 11, pp. 195-199, 2003.

[6] C. Guger, A. Schlo¨gl, C. Neuper, D. Walterspacher, T.Strein, G.Pfurtscheller, "Rapid prototyping of an EEG based brain computer interface (BCI)", IEEE Transactions on Rehabilitation Engineering. vol. 9 no. 1, pp 49–58, March 2001.

[7] G. Pfurtscheller, C. Neuper, "Motor imagery activates primary sensorimotor area in humans" , Neuroscience Letters, vol. 239, no. 2-3, pp. 65–68, 1997

[8] Wang Z, Maier A, Leopold DA, Logothetis NK, Liang H. "Single-trial evoked potential estimation using wavelets", Computers in Biology and Medicine, 37, 463 – 473, 2007

[9] M. Fatourechi1, S.G. Mason, G.E. Birch and R.K. Ward, "A wavelet-based approach for the extraction of event related potentials from EEG", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2004. vol.2 pp 737-40, 17-21 May 2004,

[10] Liu Mingyu 1, Ji Hongbing1, Zhao Chunhong, " Event Related Potentials Extraction from EEG Using Artificial Neural Network ", Proceedings of the 2008 Congress on Image and Signal Processing, Volume 01, pp 213-215, 2008

[11] Krusienski D.J., McFarland D.J. and Wolpaw J.R., "An evaluation of autoregressive spectral estimation model order for brain-computer interface applications," IEEE EMBS Ann Int Conf 2006, pp. 1323-1326, 2006.

[12] McFarland D.J., Wolpaw J.R., "Sensorimotor rhythm-based brain–computer interface (BCI): model order selection for autoregressive spectral analysis," J Neural Eng, Vol. 5, pp. 155-162, 2008.

[13] N-J Huan and R. Palaniappan, "Classification of Mental Tasks using Fixed and Adaptive Autoregressive Models of EEG Signals," Proc. of 26th Annual Intl. Conf. IEEE EMBS, CA, USA, pp 507 – 510, 2004.

[14] Hoffmann U., Vesin J.M., Ebrahimi T., "Spatial filters for the classification of event-related potentials," Proc 14th Eur Symp Artif Neural Networks, pp. 47-52, 2006.

[15] Liao X., Yao D.Z., Wu D. et al, "Combining Spatial Filters for the Classification of Single-Trial EEG in a Finger Movement Task," IEEE Trans Biomed Eng, Vol. 54, pp. 821-831, 2007.

[16] Maan M. Shaker, "EEG Waves Classifier using Wavelet Transform and Fourier Transform" , International Journal of Biological, Biomedical and Medical Sciences Vol 1, No 2, 2006.

[17] Martina Tolić, Franjo Jović, "Classification of wavelet transformed EEG signals with neural network for Imagined mental and motor tasks", Kinesiology; Vol.45 Issue 1, pp. 130-138, 2013.

[18] Vladimir Bostanov, " BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram" , IEEE Transactions on Biomedical Engineering, vol. 51, no.6, pp. 1057- 1061, June 2004.

[19] Kouhyar Tavakolian, A. M. Nasrabadi, Siamak Rezaei, "Selecting Better EEG Channels For Classification Of Mental Tasks'', IEEE International Symposium on

Circuits and System- ISCAS2004, 23-26 May 2004,

[20] Pfurtscheller G., Neuper C., Guger C. et al, "Current trends in Graz Brain-Computer Interface (BCI) research," IEEE Trans Rehabil Eng, Vol. 8, pp. 216-219, 2000.

[21] M.N.Bawane, K.M.Bhurchandi , "Classification of Mental Task Based on EEG Processing Using Self Organising Feature Map ", Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC),vol.2, pp 240-244, Aug. 2014

[22] Xiaoou Li; Xun Chen; Yuning Yan; Wenshi Wei; Wang , Z. Jane , " Classification of EEG Signals Using a Multiple Kernel Learning Support Vector Machine " Sensors 2014, Vol. 14 Issue 7 , pp.12784-12802, July2014.

[23] C. Lin and M. Hsieh, "Classification of mental task from EEG data using neural networks based on particle swarm optimization", *Neurocomputing*, vol. 72, pp 1121-1130 , Jan. 2009

[24] Emmanuel Morales-Flores, Juan Manuel Ramírez-Cortés, Pilar Gómez-Gil, Vicente Alarcón-quino, "Mental Tasks Temporal Classification Using an Architecture Based on ANFIS and Recurrent Neural Networks" , Recent Advances on Hybrid Intelligent Systems , volume 451, pp 135-146, 2013

[25] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns", Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology, vol. 115, pp 2744-2753 , Dec. 2004

[26] Hsu WY, " EEG-based motor imagery classification using neuro-fuzzy prediction and wavelet fractal features ", Journal of Neuroscience Methods, 295-302 June 2010

[27] S. Chiappa and D. Barber, " EEG classification using generative independent component analysis" Neurocomputing , 69(7-9), pp. 769-777, 2006

[28] Proakis , J.G. and Manolakis, D.G. , Digital Signal Processing: Principles, Algorithms, And Applications, 4/E , Pearson Education India , 2007

[29] Kennedy, J. and Eberhart, R.C.: 'Particle swarm optimization', in Proc. IEEE Int. Conf. Neural Networks, 1995, pp. 1942-1948.

[30] Bronzino J.D. "Principles of Electroencephalography ". In The Biomedical Engineering Handbook;CRC Press LLC: Boca Raton, FL, USA, 2000

[31] J. J. Liang, A. K. Qin, P. N. Suganthan and S. Baskar, " Comprehensive learning particle swarm optimizer for global optimization of multimodal functions ". IEEE Transactions On Evolutionary Computation, Vol. 10, No. 3, June 2006 281–295.

[32] M. Clerc, J. Kennedy, " The particle swarm-explosion , stability and convergence in a multidimensional complex space" , IEEE Transactions on Evolutionary Computation 6 (1) (2002).

[33] BCI Competition III Data Set V Mental Imagery, Multi-Class, , 2004[Online]. Available: http://www.bbci.de/ competition/iii/desc_V.html

[34] C. W. Therrien, " Discrete Random Signals and Statistical Signal Processing ", New Jersey: Prentice-Hall,2004

[35] Erman Acar, 2011. "Classification of motor imagery tasks in EEG signal and Its application to a brain-computer interface for Controlling assistive environmental devices", MS thesis, Middle East Technical University,Turkey, February 2011

[36] R. Aler, I. Galván, and J. Valls, "Evolving spatial and frequency selection filters for brain-computer interfaces," in 2010 IEEE World Congress Comput. Intell. WCCI 2010, Barcelona, Spain, pp. 1–7, 2010

[37] S.R. Kheradpisheh, A. Nowzari-Dalini, R. Ebrahimpour, M. Ganjtabesh, " An evidence-based combining classifier for brain signal analysis", PloS one 9 (1) (2014) e84341