ISSN 2047-3338

A Neural Network Approach for EEG Classification in BCI

G.V. Sridhar[1] and Dr. P. Mallikarjuna Rao[2]

[1]ECE, Al-Ameer College of Engineering & IT, Visakhapatnam, A.P, India

[2]ECE, A.U. College of Engineering, Visakhapatnam, A.P, India

[1]sridhar1209@gmail.com, [2]pmraoauece@yahoo.com

***Abstract*—A Brain Computer Interface (BCI) is a new communication channel allows a person to control special computer applications like a computer cursor or robotic limb through the use of his/her thoughts. BCIs had become an active research area in the last decade. BCI research is based on recording and analyzing electroencephalographic (EEG) data and recognizing EEG patterns associated with various mental states. Supervised classification methods are employed to recognize these EEG activity patterns to learn the mapping between the EEG data and classes corresponding to mental tasks. Selecting an optimal frequency band and extracting a good set of features is still an open research problem. In this paper, it is proposed to investigate EEG signals, extract features of motor imagery in the frequency domain using Hilbert transform, to compute the maximum and minimum energies. For efficient classification, Principal Component Analysis is used for feature reduction. Classification is carried out using multilayer perceptron with different learning rate and Momentum.**

***Index Terms*— Brain Computer Interface (BCI), EEG, Hilbert Transforms, Principal Component Analysis (PCA) and Multilayer Perceptron (MLP)**

I. Introduction

A Brain Computer Interface (BCI) allows a person to control special computer applications like a computer cursor or robotic limb through the use of his/her thoughts. The idea was provision of a new communication channel to those who were paralyzed but were cognitively intact like those suffering from the so called locked-in syndrome. BCIs had become an active research area in the last one decade as science grew rapidly [1], [2]. BCI research is mainly based on recording and analyzing electroencephalographic (EEG) brain activity and recognizing EEG patterns associated with various mental states. For example, imagining a movement of the right hand is represented with an EEG activity pattern on the left of the motor cortex. Other frequent mental tasks included movements of the left hand, of the toes and the tongue. Mental tasks were carefully chosen to activate different parts of the brain, which made detection easier.

The increasing success of BCI systems was partially due to increased understanding of the brain oscillation dynamics that generated EEG signals. Sensorimotor activity like body movements or mental imagery including imagined body movement changes the oscillatory patterns leading to amplitude suppression called event related desynchronization. The amplitude enhancement known as event related synchronization is recorded as the Rolandic mu rhythm (7-13 Hz) and the central beta rhythms above 13 Hz. Supervised classification methods are employed to recognize these EEG activity patterns to learn the mapping between the EEG data and classes corresponding to mental tasks like movement of the left hand [3].

From data mining point of view this was difficult because of two reasons. First, EEG data are noisy and correlated as many electrodes were required to be fixed on a small scalp surface with each electrode measuring the activity of thousands of neurons [4], [5]. Selecting an optimal frequency band and extracting a good set of features was still an open research problem. The data quality is also affected by the different degrees of the subject's attention and changes in concentration. Traditionally, classical linear classifiers like Fisher's linear discriminant were favored [3], [6], [7]. More recently, a variety of machine learning classifiers were applied including neural networks like multi-layer perceptrons [8], [9], probabilistic classifiers [10], lazy learning classifiers including k-nearest neighbor [6] and state of the art classifiers like support vector machines [4]. Lotte et al., noted that some classical classification algorithms like decision trees and ensembles of classifiers were yet to be evaluated. Thus, the goal of this study is to evaluate BCI classification task using multilayer perceptron, based on a benchmark dataset, and using preprocessing and feature selection methods.

II. Methodology

To investigate the proposed method publicly available dataset available in [11] was obtained. Labview was used to implement the Hilbert Transform for feature extraction. The maximum and minimum energy are computed for all the evoked responses. Principal Component Analysis is used for feature reduction. Classification is carried out using multilayer perceptron with different learning rate and Momentum.

A. Dataset

The IV A dataset used in the brain computer interface competition provided by Intelligent Data Analysis Group is used as dataset for experimentation [11]. The EEG recordings complied from five healthy subjects sitting in a chair with arms resting on armrests forms the dataset. Visual cues for 3.5 s were shown for the subject to perform 3 motor imageries: (L) *left* hand, (R) *right* hand, (F) right *foot*. The dataset contains continuous signals of 118 EEG channels and markers that indicate the time points of 280 cues for each of the 5 subjects (*aa*, *al*, *av*, *aw*, *ay*). Subject *aa* was used in our study.

B. Hilbert Transforms

Hilbert transforms play an important role in signal processing and is used to extract feature from the EEG signals in this study. Analytic signal, bandpass sampling, minimum phase networks, and spectral analysis are based on Hilbert transform relationships.

The Hilbert transform [12] of a function $x(t)$ is given by:

$$h(t) = H[x(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau$$

Using the Fourier identities, the Fourier transform of the Hilbert transform of $x(t)$ id

$$h(t) \Leftrightarrow H(f) = -j\,\mathrm{sgn}(f)\,X(f)$$

where $x(t) \Leftrightarrow X(f)$ is a Fourier transform pair and

$$\mathrm{sgn}(f) = \begin{cases} 1 & f > 0 \\ 0 & f = 0 \\ -1 & f < 0 \end{cases}$$

The artifacts in the obtained frequencies using Hilbert transform are removed using a bandpass Chebyshev filter [13] such that all frequencies below 5 Hz and above 20Hz are eliminated.

C. Principal Component Analysis (PCA)

Feature reduction refers to multidimensional space mapping into lower dimension space. These techniques are usually a pre-process to machine learning and statistics tasks include that of prediction and pattern recognition. The feature space contributes to classification, cuts pre-processing costs and lowers effects of the classification ‘peaking phenomenon.’ This in turn greatly improves overall performance of classifier. Principal component analysis (PCA) is a data analyzing technique used to compress high dimensional vector data sets into low dimensional ones. PCA is derived from many starting points and optimization criteria, the most important of them being minimization of mean-square error in data compression, locating mutual orthogonal directions in data with maximal variances, and data de-correlation using orthogonal transformations.

PCA is a common technique for locating finding patterns in high dimension data. Statistics as a subject is based on the idea that you have a big dataset which you want to analyze the set terms of relationships between individual points in that set. PCA’s goal is data dimensionality reduction while retaining much of variation present in the original data set. It is a method of identifying data patterns and expressing it in a way to highlight both similarities and differences [14]. The following figure 1 shows the PCA algorithm flow.

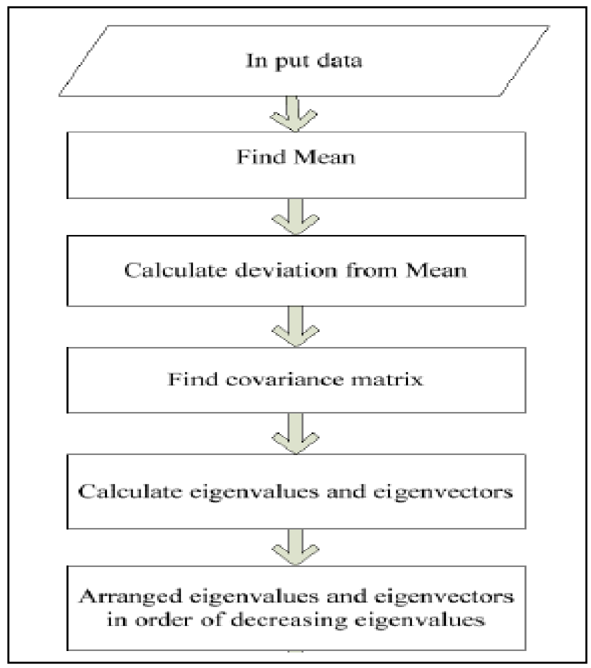

Fig. 1: PCA algorithm flow

D. Multilayer Perceptron (MLP)

A human brain comprises of about ten billion neurons of high structural and functional complexity. They are densely interconnected resulting in a complex architecture and an intelligence level not achieved by any artificial system. Several mathematical models were developed to represent neurons and their interconnections. Artificial neural networks (ANN) were seen as attempts to reproduce human brain potentialities, specially its learning ability. The first neuron mathematical model proposed by McCulloch and Pitts, had a binary output and inputs, each of different excitatory or inhibitory gains, known as synaptic weights (or weights). The values of input signals and related weights determined neuron output. Hence, perceptron or artificial neuron is a mathematical model of a neuron cell and the basic unit compounding artificial neural network. Perceptron architecture includes a set of *n* inputs (*xi*), each related to a weight (*wi*) and an activation function (*fi*). Perceptrons are organized to form layers, where all are linked to same inputs but with distinct outputs. Such a network is called a perceptron network. Perceptron networks achieve good performances only when recognizable pattern is linearly separable; hence, they cannot not solve complex classification problems having non-linearly separable patterns, instead multilayer perceptron networks should be used.

Multilayer perceptron (MLP) networks have an input layer (*Xi*), one or more intermediary or hidden layers (*HL*) and an output layer (*Y*). A weight matrix (*W*) is defined for each layer. The ANN topology solves classification problems with non-linearly separable patterns and is also used as a universal function generator.

MLP have two clear phases: training and execution. With this network topology, it is impossible to use the delta rule directly for training, as the rule does not permit weight recalculation for subterranean layers. Hence, a popular algorithm for MLP network training is backpropagation with variants. This learning approach is more complex than that for a perceptron network and it is of the supervised type [15].

The basic MLP learning algorithm is developed below [16].

1. Initialise network, with weights set to random numbers between -1 and +1.
2. Present first training pattern, and obtain output.
3. Compare network output with target output.
4. Propagate error backwards.

(a) Correct output weights layer with the formula below:

$$w_{ho} = w_{ho} + \eta \delta_o o_h$$

where w_{ho} is the weight connecting hidden unit h with output unit o, η is the learning rate, o_h is the output at hidden unit h. δ_o is given as follows:

$$\delta_o = o_o(1 - o_o)(t_o - o_o)$$

Where o_o is output at node o of output layer, and t_o is target output for that node.

(b) Correct input weights using following formula.

$$w_{ih} = w_{ih} + \eta \delta_h o_i$$

where w_{ih} is weight connecting node i of input layer with node h of hidden layer, o_i is input at node i of input layer, δ_h is calculated as follows.

$$\delta_h = o_h(1 - o_h)\sum_o (\delta_o w_{ho})$$

5. Calculate error, by taking average difference between target and output vector. The following function can be used as an example.

$$E = \frac{\sqrt{\sum_{n=1}^{p} (t_o - o_o)^2}}{p}$$

Where p is number of units in output layer.

6. Repeat from 2 for each pattern in training set to complete one epoch.
7. Shuffle training set randomly, to prevent network being influenced by the data order.
8. Repeat from step 2 for a specific number of epochs, or till the error ceases to change.

The transfer function is selected by the designer after which parameters and they will be adjusted by a learning rule to ensure that the neuron input/output relationship meets a specific goal. Sigmoid and tanh transfer function are used in this study. The transfer function is obtained as follows:

Sigmoid function

$$P(t) = \frac{1}{1+e^{-t}}$$

Tanh function

$$\tanh = \frac{e^{2x}-1}{e^{2x}+1}$$

III. Related works

Some of the works in literature explored the use of Hilbert Transforms for extraction of features from the EEG signals. Huang, et al., [17] investigated wavelet transform and Hilbert-Huang transforms (HHT) methods for processing EEG signal. The experimental results showed that main features of the EEG are extracted efficiently in both the methods, however the HHT are more accurate when expressing the EEG distribution in time and frequency domain. Lei Wang, et al., [18] proposed extracting features from EEG data based on motor imagery using Hilbert Huang transform (HHT). In the proposed method, HHT with genetic algorithm (GA) is used for selection of the most relevant features from the frequency domain. Experimental results show that HHT and GA achieve much higher classification accuracy when compared with traditional frequency feature extraction methods.

Dias, et al., [19] proposed a new PCA as a Variable Subset Selection (VSS) method to produce a ranked list of original variables. The variables were ranked according to its ability to discriminate between tasks. The aim of the proposed approach was to find few relevant variables for discrimination in a high dimensional variable space. The experimental results showed that at least 83% relevant variables were selected for optimal subsets and 100 % of predominant variables were selected for all optimal subsets.

IV. Results and discussion

The EEG signals are preprocessed using Hilbert transform for feature extraction and PCA for feature reduction. Experiments are conducted using ten fold cross validation using a MLP classifier. Table 1 gives the classifier model parameters. Experiments are conducted using sigmoid and tanh functions and learning rate varying from 0.1 to 0.4 and momentum of 0.1 to 0.3.

TABLE 1: MLP PARAMETERS

Random Number Seed	0
Learning Rate	0.1 to 0.4
Learning Rate Function	Static learning rate
Constant Bias Input	1
Training Iterations	500
Training Mode	Batch Training - weight changes are applied at the end of each epoch
Transfer Function	Sigmoid, S-shape function between +1 and 0 ; Tanh, S-shape function between +1 and -1
Momentum	0.0.1 to 0.3
Weight Decay	0
Bias Input Value	1
Num Inputs:	49
Hidden Layer 1	15
Output Layer	2
Total Neurons	17

The classification accuracy obtained from the proposed method using MLP as the classifier is shown in Fig. 2. Table I tabulates the classification accuracy. Table 2 gives the classification accuracy. It is seen that tanh transfer function performance is the lowest.

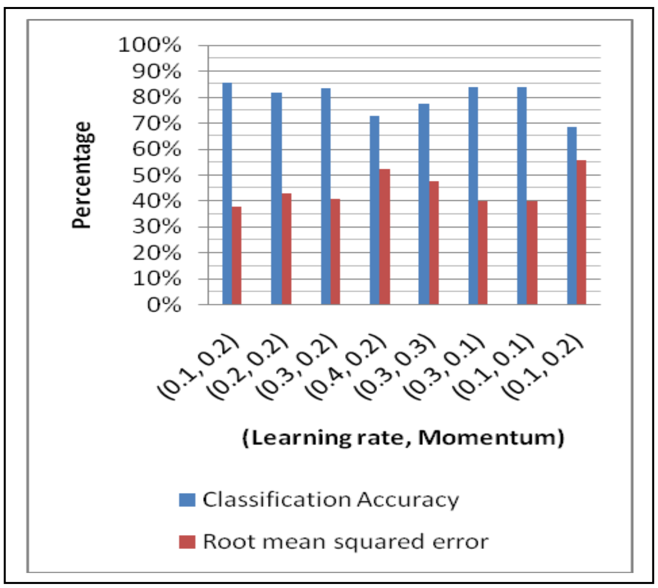

Fig. 2: Classification accuracy and RMSE for different learning rate and momentum

TABLE 2: CLASSIFICATION ACCURACY FOR DIFFERENT LEARNING RATE AND MOMENTUM

Transfer function	Learning rate	Momentum	Classification Accuracy	Root mean squared error
sigmoid	0.1	0.2	**85.71%**	0.378
sigmoid	0.2	0.2	81.55%	0.4296
sigmoid	0.3	0.2	83.33%	0.4082
sigmoid	0.4	0.2	72.62%	0.5233
sigmoid	0.3	0.3	77.38%	0.4756
sigmoid	0.3	0.1	83.93%	0.4009
sigmoid	0.1	0.1	83.93%	0.4009
tanh	0.1	0.2	68.45%	0.558

Fig. 3 and Fig. 4 show the Precision, recall and f Measure.

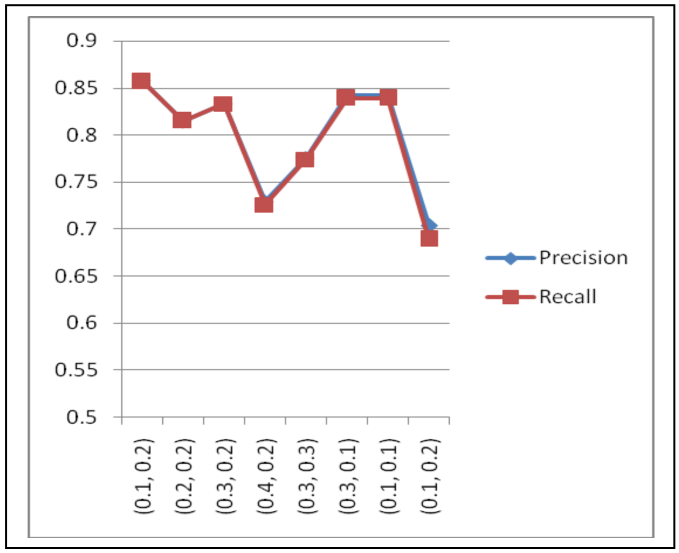

Fig. 3: Precision and Recall

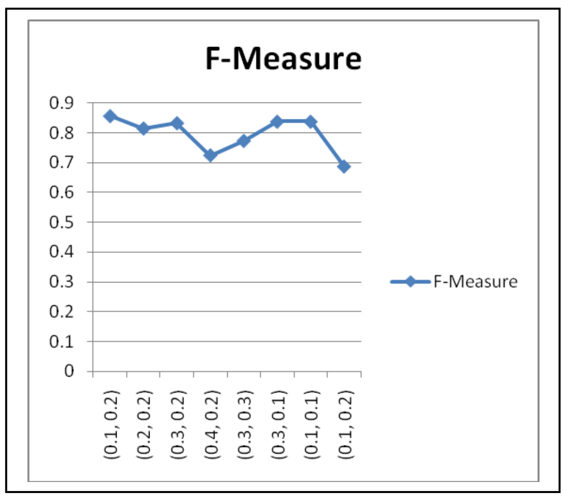

Fig. 4: f Measure for different learning rate and momentum of MLP

V. CONCLUSION

In this paper, it was proposed to extract features from EEG data by converting the time series EEG data to frequency domain using Hilbert Transform and PCA for feature reduction. The preprocessed signal is classified using Multilayer Perceptron using sigmoid and tanh function. Experiments are conducted using ten fold cross validation and different learning rates and momentum. The accuracy obtained is comparable with the results obtained from other researchers in literature. The proposed method is extremely fast in both feature extraction and classification. Further work needs to be done to improve the classification accuracy.

REFERENCES

[1] Wolpaw, J.R., Birbaumer, N, Heetderks, W.J., McFarland, D.J, Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A, Robinson, C.J., Vaughan, T.M. (2000): Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering* 8:164-173.

[2] Dornhege, G., Millán, J. del R., Hinterberger, T., McFarland, D.J. and Müller K.-R. (2007): *Toward BrainComputer Interfacing*. Cambridge, MIT Press.

[3] Lotte, F., Congedo M., Lecuyer A., Lamarche F. And Arnalsi B. (2007): a review of classification algorithms for EEG-based brain-computer interfaces. Journal of Neural Engineering 4: R1-R13.

[4] Lee, F., Scherer, R., Leeb, R., Neuper, C., Bischof, H. and Pfurtscheller, G. (2005): A comparative analysis of multi-class EEG classification for brain computer interface. *Proc. 10th Computer Vision Winter Workshop (CVWW)*, Technical University of Graz, Austria.

[5] Thulasidas, M., Guan C. and Wu, J. (2006): Robust classification of EEG signal for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14:24-9.

[6] Blankertz B., Curio, G. and Müller K.-R. (2001): Classifying single trial EEG: towards brain computer interfacing. In *Advances in Neural Information Processing Systems*, T.G. Diettrich, S. Becker and Z. Ghahramani, ed., 14: 157-164.

[7] Müller, K-R., Anderson C.W and Birch G.E (2003): Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11(2): 165-169.

[8] Anderson C.W. and Sijercic Z. (1996): Classification of EEG signals from four subjects during five mental tasks. *Proc. International Conference on Engineering Applications of Neural Networks (EANN).*

[9] Kubat, M., Koprinska, I. and Pfurtcheller G. (1998). Learning to classify biomedical signals. In *Machine Learning and Data Mining: Methods and Applications*, R.S. Michalsi, M. Kubat and I. Bratko (ed.), Wiley.

[10] Barreto, G.A., Frota R.A. and Medeiros F.N.S. (2004): On the classification of mental tasks: a performance comparison of neural and statistical approaches. *Proc. IEEE Workshop on Machine Learning for Signal Processing.*

[11] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. IEEE Trans. Biomed. Eng., 51(6):993-1002, June 2004.

[12] Pei, S.-C.; Jaw, S.-B.; "Computation of discrete Hilbert transform through fast Hartley transform,"*Circuits and Systems, IEEE Transactions on*, vol.36, no.9, pp.1251-1252, Sep 1989.

[13] Steven. W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing" 1997.

[14] Iftikhar Ahmad , Azween Abdullah, Abdullah Alghamdi, Muhammad Hussain, "Optimized intrusion detection mechanism using soft computing techniques",© Springer Science+Business Media, LLC 2011

[15] Victor Hugo Costa Albuquerque, Auzuir Ripardo de Alexandria, Paulo César Cortez, João Manuel R. S. Tavares," Evaluation of Multilayer Perceptron and Self-Organizing Map Neural Network Topologies applied on Microstructure Segmentation from Metallographic Images"

[16] Leonardo Noriega ,"Multilayer Perceptron Tutorial"

[17] Manling Huang, Pingdong Wu, Ying Liu, Luzheng Bi, Hongwei Chen, "Application and Contrast in Brain-Computer Interface between Hilbert-Huang Transform and Wavelet Transform," icycs, pp.1706-1710, 2008 The 9th International Conference for Young Computer Scientists, 2008.

[18] Lei Wang; Guizhi Xu; Jiang Wang; Shuo Yang; Weili Yan; "Motor Imagery BCI Research Based on Hilbert-Huang Transform and Genetic Algorithm", Bioinformatics and Biomedical Engineering, May 2011 5th international conference.

[19] S.Dias, M.Kamrunnahar, P.M.Mendes, S.J.Schiff, & J.H.Correia; variable subset selection for Brain-Computer interface-PCA-based Dimensionality Reduction and Feature Selection; BIOSIGNALS 2009 - International Conference on Bio-inspired Systems and Signal Processing