Function Classification of EEG Signals Based on ANN

Rajesh Singla, Neha Sharma

***Abstract*—The motor imagery is limited in pattern variety, so in our work, six motor imageries including wrist, elbow, wrist rotation clockwise/anticlockwise and ankle backward/forward moment were used in this system. This paper described the auditory paradigm for recording of motor imagery signals and the relevant coefficient was used for signal analysis and recognition. EEG signals were decomposed into wavelet coefficients by discrete wavelet transform on which SVD technique is applied to get singular value used as feature vectors, presenting them into ANN classifier.**

***Keywords*— BCI, EEG, Wavelet Transform, SVD,ANN**

I. INTRODUCTION

A Brain-Computer Interface (BCI) is a communication system capable of transforming the person's cognitive functions into control commands that let the user interact with external devices [1], [2]. The basic operation of a BCI is to record the cerebral bioelectric activity through electrodes in order to differentiate between several mental tasks. This kind of systems creates a natural way of human-machine communication because they translate intentions into orders to interact with the environment without performing any physical movement. Thus, the BCI systems are of great interest to people with severe disabilities or mobility limitations. They can improve their quality of life and assist them in various daily tasks.

Motor-Imagery based BCI (MI-BCI) is one that decodes the brain signals from the imagination of left or right hand movement and is developed based on the studies that showed neural activation for movement as well as movement imaginations. The movement related features of EEG are derived from Movement Related Potential (MRP) or Event-related Desynchronization/ Synchronization (ERD/ ERS). MRP based methods carry out time domain analysis of EEG data [3]. The neuronal activation during the preparation of movement results in contra-lateral negativation of EEG and is termed as readiness potential (RP). Several BCI systems employ features related to RP to discriminate left or right movements. MRP analysis requires simple pre-processing steps such as denoising and filtering. ERD/ERS [20] features are the commonly used neural features in EEG based BCIs. Movement preparation is accompanied by a contra lateral power decrease in mu rhythm termed as ERD. Movement and post-movement relaxation causes an increase in beta rhythm termed as ERS. The ERD/ERS analysis requires precise time and frequency patterns. The non-stationary behavior of EEG signals is a major issue in these types of analysis.

In the literature of motor imagery classification, several feature extraction methods were used to extract the most informative set of features, representing them as input patterns into classifiers. An autoregressive (AR) model was applied to short overlapping EEG segments. From the AR spectrum, band power is calculated in several frequency bands and the power sum is used as independent variables in a linear function that defines the control signals [4]. A number of feature extraction methods exist for BCI applications, such as Fast Fourier Transformation (FFT) [18], Principal Component Analysis (PCA) [5] and Independent Component Analysis (ICA) [19].. Also, CSP and adaptive autoregressive parameters (AAR) were used as feature extraction methods in [7]. Recently, algorithms using weighted time–frequency analysis [8] were developed.

Wavelet Transform (WT) was proposed to address the problem of poor temporal resolution in non stationary EEG signals. Wavelet coefficients were used as feature vectors identifying characteristics of the signal that were not apparent from the original time domain signal [9]. Supervised classification methods are employed to recognize the patterns of EEG activities. Several methods have been proposed to BCI. As the classification methods, Linear discriminant analysis (LDA) and fisher's linear discriminant analysis (FLDA) [12], support vector machine (SVM) [10] lazy learning classifiers such as knearest neighbor [11], artificial neural networks (ANNs) were investigated to design EEG-based BCI systems [13].

Due to having non-stationary signals, poor signal to noise ratio, highly overlapped classes, small sample size and high dimensional feature sets, EEG classification can be categorized into complex problems . Combining classifiers is an approach to improve the performance in classification particularly for complex problems [15].

The paper has been organized as follows: Section 2 introduces the experiment datasets including the collection of data and data preprocessing. Section 3 describes the basic theory of the Discrete Wavelet Transform (DWT) along with SVD and the structure of the neural networks (NN). In Section 4, the Matlab based computation process is described briefly. Section 5 makes the conclusions to summarize the study.

Manuscript received January 15, 2014.

Rajesh Singla,is with the Instrumentation and Control Engineering Department, National Institute of Technology Jalandhar, PIN-144011, India

Neha Sharma, is with the Instrumentation and Control Engineering Department, National Institute of Technology Jalandhar, PIN-144011, India

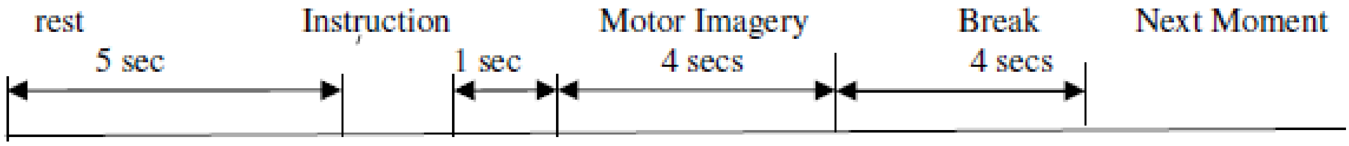

Figure 2. The auditory paradigm of experiment

II. EXPERIMENTAL DATASETS

A. *Setup Used*

Data is acquired with the help of Supersac software The EEG signal is acquired using RMS EEG 32 Super Spec system. The Super Spec ensures high-resolution, authentic data acquisition through its software and Head Box. The standard parts of the EEG hardware include adaptor box, head box, connecting cable and PC.

The Head Box is used for connecting electrodes from the scalp to the hardware unit. The signal generated is amplified and then sent to adaptor Box for signal conditioning. The digital signal generated then, passes to the PC where it is displayed on the screen on Super Spec software designed for display of EEG Signals. The leads connected to the Head Box. The Head Box minimizes the noise pickups. The total integration of analog and digital processing in the compact Head Box gives excellent signal to noise ratio too.

B. *Electrode Placement*

Three Ag/Agcl circular electrodes were mounted on the C3, Cz, and C4 with the references at the both mastoids and the ground at the Fpz over the forehead according to the 10-20 system (Fig.1,3 Superspec was used to record EEG signals with the resolution of 0.5μV and the time constant of 0.3s. The sampling frequency was 256Hz with 60Hz upper cut-off filtering and lower cut-off frequency of 0.1Hz including a notch filtering at 50Hz. The reluctance was keeping below 5K during detection.

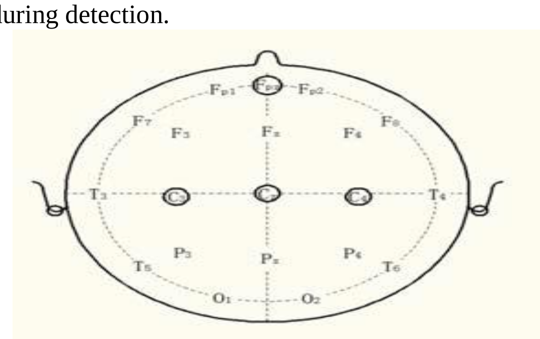

Fig.1 Electrode placement map

Figure 3. The EEG electrodes for data recording

C. *Post-Historic Training*

All the subjects were trained before the recording of actual data. At the starting of training, a video clip lasting for 3 minutes containing the 3 samples of each motor imagery moment was shown to the subjects who explained them how to perform the task.

After the video clip each subject was trained for all the moments displayed in the video. Each session contained 3 moments of each movement and lasted for 3 minutes. There were about 5-6 sessions of training were given to persons which depend on their state and bearing capability.

The participants were sitting in an armchair, faced a speaker 2 meters in the distance, and were told about the whole experiment paradigm for comprehension by an introduction audio with a demo [16,17]. Participants had to perform six types motor imagery according to the instruction which they heard including left/right hand and feet movements.

D. *Auditory Paradigm*

The participants were sitting in an armchair, faced a speaker 2 meters in the distance, and were told about the whole experiment paradigm for comprehension by an introduction audio with a demo. Participants had to perform six types motor imagery according to the instruction which they heard including wrist movement, left/right rotation of wrist, elbow movement and forward/backward ankle movements. At the beginning of a trial, an attention voice sounded. After 6s a voice instruction to notice the participants what imagery should be performed in the next. After a pause of 2s, a start ringed and participants should begin or already to perform the specified motor imagery and kept this step for 4 seconds. Then a stop voice told the participants the imagery had finished, after that there was a break lasting between 4s before the next trial began. One round consisted of 3 trials for each motor imagery task, a total of 18. Three rounds with a short rest between each one were carried out in this part, obtaining 54 trials in total

III. FEATURE EXTRACTION AND CLASSIFICATION

A. **Feature Extraction**

1) *DWT:* Wavelet transform decomposes a signal into a set of basis functions These basis functions are called wavelets.

Wavelets are obtained from a single prototype wavelet y(t) called mother wavelet by dilations and shifting

$$\psi_{a,b}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \qquad (1)$$

where a is the scaling parameter and b is the shifting parameter

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-3 Issue-6, January 2014

DWT gives precise time-frequency information about the signal. It decomposes the signal in a number of sub-bands at different scales, according to the number of decomposition levels. DWT is defined by two functions referred as scaling and detail function

$$\psi(t)=\sqrt{2}\sum_{l\in\mathbb{Z}} h_o\, \psi(2t-l), \quad (2)$$

$$\psi(t)=\sqrt{2}\sum_{l\in\mathbb{Z}} h_1\, \psi(2t-l), \quad (3)$$

The frequency responses of equation 2 and 3 correspond to low pass and high pass FIR filters, respectively; h_o are the low-pass filter coefficients and h_1 are the high-pass filter coefficients related to a chosen wavelet . From these equations a signal can be estimated as:

$$x(t)=\sum_{l\in\mathbb{Z}} c_{j,k}\ \psi_{j,k}(t)+\sum_{j=J}\sum_{k\in\mathbb{Z}}^{\infty} d_{j,k}\ \psi_{j,k}(t) \quad (4)$$

where

$$c_{j,k}=\sum_m h_o\ c_{j-1,2k-l}$$

$$d_{j,k}=\sum_m h_1\ c_{j-1,2k-l}$$

Approximation and details coefficients are obtained through equations 2 and 3, which describe a set of recursive quadrature mirror filters. The so-called sub-band coding algorithm is illustrated in Fig. 4 with a two wavelet decomposition levels. The frequency ranges corresponding to each sub-band depends on the sampling frequency of the signal.

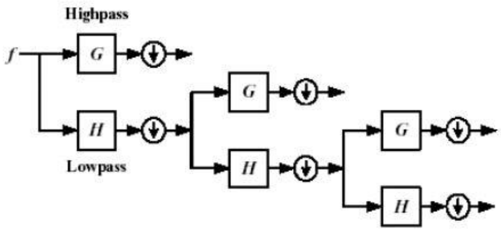

Fig.4 Wavelet Decomposition

In practice, such transformation will be applied recursively on the low-pass series until the desired number of iterations is reached.

The DWT computes "wavelet coefficients" To obtain the informative set of features, all EEG segments were decomposed into wavelet coefficients using discrete wavelet transform levels was chosen to be 4. Also, by trial and error, the smoothing feature of the Daubechies wavelet of order 3 (db2) was used as wavelet function.

2) *Dimensionality Reduction:* Singular value decomposition (SVD) can be looked at from three mutually compatible points of view[15]. On the one hand, we can see it as a method for transforming correlated variables into a set of uncorrelated ones that better expose the various relationships among the original data items. At the same time, SVD is a method for identifying and ordering the dimensions along which data points exhibit the most variation. This ties in to the third way of viewing SVD, which is that once we have identified where the most variation is, it's possible to find the best approximation of the original data points using fewer dimensions. Hence, SVD can be seen as a method for data reduction.

As an illustration of these ideas, consider the 2-dimensional data points in Figure 5. The regression line running through them shows the best approximation of the original data with a 1-dimensional object (a line). It is the best approximation in the sense that it is the line that minimizes the distance between each original point and the line. If we drew a perpendicular line from each point to the regression line, and took the intersection of those lines as the approximation of the original data point, we would have a reduced representation of the original data that captures as much of the original variation as possible. Notice that there is a second regression line, perpendicular to the first, shown in Figure 5. This line captures as much of the variation as possible along the second dimension of the original data set. It does a poorer job of approximating the original data because it corresponds to a dimension exhibiting less variation to begin with. It is possible to use these regression lines to generate a set of uncorrelated data points that will show sub-groupings in the original data not necessarily visible at first glance. SVD is based on a theorem from linear algebra which says that a rectangular matrix *A* can be broken down into the product of three matrices - an orthogonal matrix *U*, a diagonal matrix *S*, and the transpose of an orthogonal matrix *V* . The theorem is usually presented something like this:

$$A_{mm}= U_{mm}S_{mm}V_{nn}^{t} \quad (5)$$

where $U^TU = I$, $V^TV = I$; the columns of *U* are orthonormal eigenvectors of AA^T , the columns of *V* are orthonormal eigenvectors of A^TA, and *S* is a diagonal matrix containing the square roots of Eigen values from *U* or *V* in descending order.

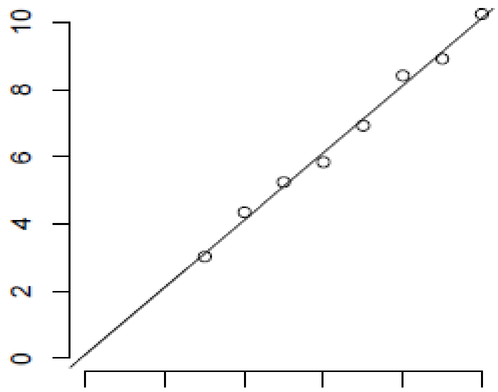

Figure 5: Best-fit line reduces data from two dimensions into one.

The basic idea behind SVD is taking a high dimensional, highly variable set of data points and reducing it to a lower dimensional space that exposes the substructure of the original data more clearly and orders it from most variation to the least.

What makes SVD practical for NLP applications is that one can simply ignore variation below a particular thresh hold to massively reduce ones data but be assured that the main relationships of interest have been preserved.

B.Function Classification: Artificial neural network (ANN) is a mathematical model that mimics some functional aspects of a biological neuron network [13]. The ANN consists of an interconnected group of artificial neurons. These neurons are basic computational elements, often called either nodes or units. The node receives input from some other nodes or from an external source. Each input has an associated weight w, which can be modified so as to model synaptic learning. The node computes some function f of the weighted sum of its inputs:

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$y_i = f\left(\sum_{j=0}^{n-1} w_{i,j}\, x_j\right) \quad (6)$$

where *x* is input vector, *w* is weight vector.

Single neuron is a relatively simple computational element so they have to be connected into network in order to complete complex tasks. Two or more neurons can be combined in a layer, and a particular network could contain one or more such layers. Each output of one layer is interconnected to input of next layer through the weight matrix W. Usually ANN has one input layer, one output layer and one or more hidden neuron layers. Theoretically a single hidden layer network can solve task of any complexity [13].

Training a neural network is a numerical optimization of a nonlinear error function. There is no single best method for this problem. Method must be chosen based on the characteristics of the problem to be solved and network configuration. For feed forward artificial neural networks with the most popular differentiable activation functions and error functions, three general types of algorithms have been found to be effective for most practical purposes [13]:

For classification we use a pattern recognition network, which is a feed-forward network with tan-sigmoid transfer functions in both the hidden layer and the output layer and there are 20 neurons in each hidden layer. The network has six output neurons, because there are six categories associated with each input vector. Each output neuron represents a category. When an input vector of the appropriate category is applied to the network, the corresponding neuron should produce a 1, and the other neurons should output a 0.

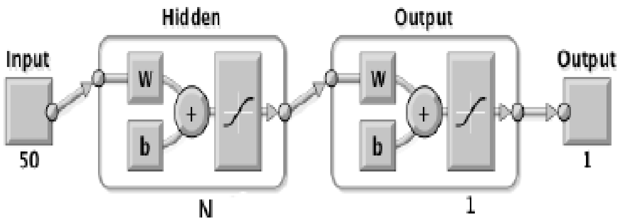

Fig.6 Neural Network Configuration

The pattern recognition network uses the default Scaled Conjugate Gradient algorithm for training. The application randomly divides the input vectors and target vectors into three sets. 60% are used for training. 15% are used to validate that the network is generalizing and to stop training before over fitting The last 20% are used as a completely independent test of network generalization.

IV. METHOD

In the present study, EEG signals were recorded during motor imagery task from ten subjects. In each trial, subjects were asked to imagine one of predefined tasks (ankle backward, ankle forward, Wrist , elbow, wrist clockwise and wrist anticlockwise movement) in the training phase which is from t= 8s to t= 12s (see figure 2). Using super spec software by placing the electrodes at C3,C4 and Cz location. Since these channels are placed in the motor cortex area, these channels can be used to recognize the changes of ongoing EEG pattern during motor imagery tasks better than other ones.

Training dataset consists of 72 labeled trials. First 12 trials belong to class 1, next 12 belongs to class 2 and so on upto class 6. Testing dataset consists of 18 trials. Each trial consists of 1024 samples from each of 9 channels. The sampling rate of 256 Hz and the recording length is 4s for each trial.

The raw signal is in time domain and is random or stochastic in nature, this means no data can be excluded from classification using ANN. Fig. 4 shows discrete wavelet transform of EEG data. The resulting signal in frequency domain is less complex, energy is compressed into first few coefficients and all others are relatively small. These small coefficients can be omitted from classification.

DWT data plot in Fig. 4 shows that higher coefficients are less than our predefined value. In this experiment, we chose 7^{th} node for classification, ignoring the rest of DWT data. Since the signal has 1024 samples, we are using less than 13% of original dataset size i.e 128 samples for each trial. Also, this method takes advantage of DWT feature to concentrate low frequency components into first coefficients. This means that there is no need of using any filters for high frequency noises.

In this experiment, all computations were carried out using MATLAB. The method of DWT was applied to convert the time domain data to frequency domain. Then 128 coefficients of every trial were taken for further investigation. This procedure effectively compressed data of the original set.

Further, SVD compressed the data to one single value for each motor imagery moment which are then given to ANN classifier. For our analysis we use 90 trials for training data and 24 trials for testing.

Classification of data was performed using the MATLAB Neural Network Toolbox. For experiments a several feed forward networks with one hidden layer was created. Number of hidden neurons was 10, 20, 30, 40, 50 and 70. In all cases output layer has only six neurons with a tansigmoid transfer function. In Fig. 5 a general network configuration is shown, where N is the number of hidden neurons.

V. RESUTS AND DISCUSSION

The classification results are shown in fig. 9 using confusion matrix in which the diagonal cells in each table show the number of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red).

The overall accuracy is 82%, the accuracy corresponding to each class is shown by red block in . The results for all three data sets (training, validation, and testing) show very good recognition.

The ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (1 - specificity) as the threshold is varied. A perfect test would show points in the upper-left corner as shown in fig 8.

The regression plot for testing, validation and testing data is shown is fig 7 which depicts the relationship b/w the output and input data.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

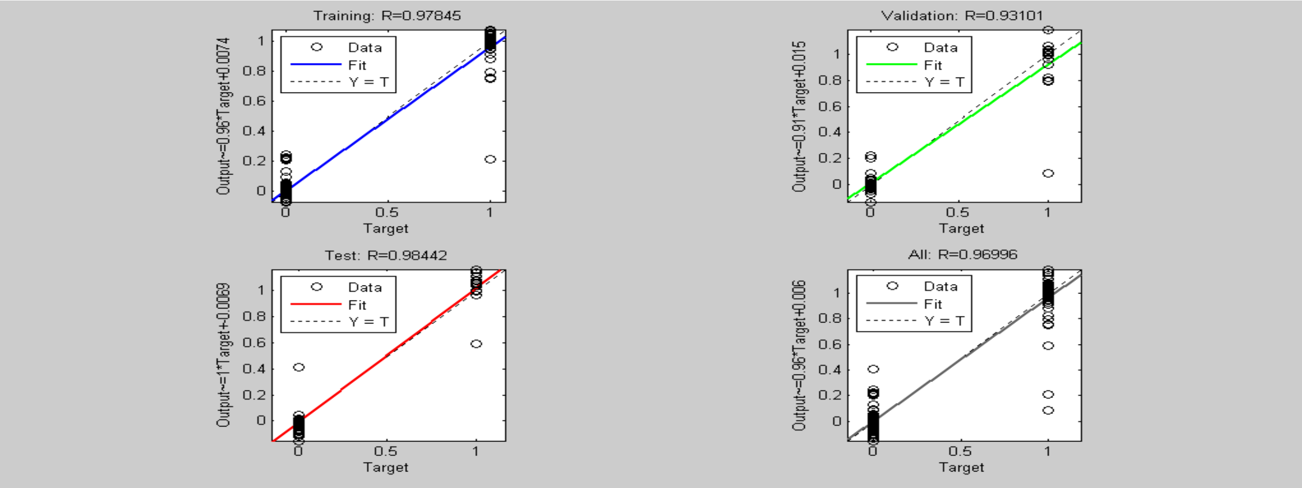

Fig.7- Regression plot for training, validation and testing data

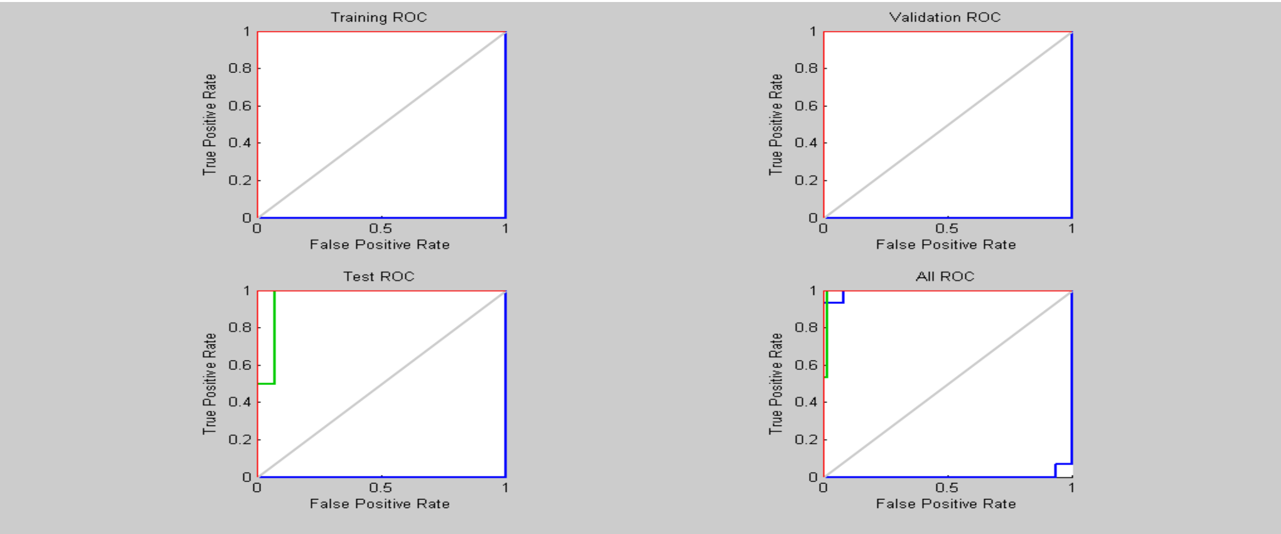

Fig.8- ROC for training, validation and testing data

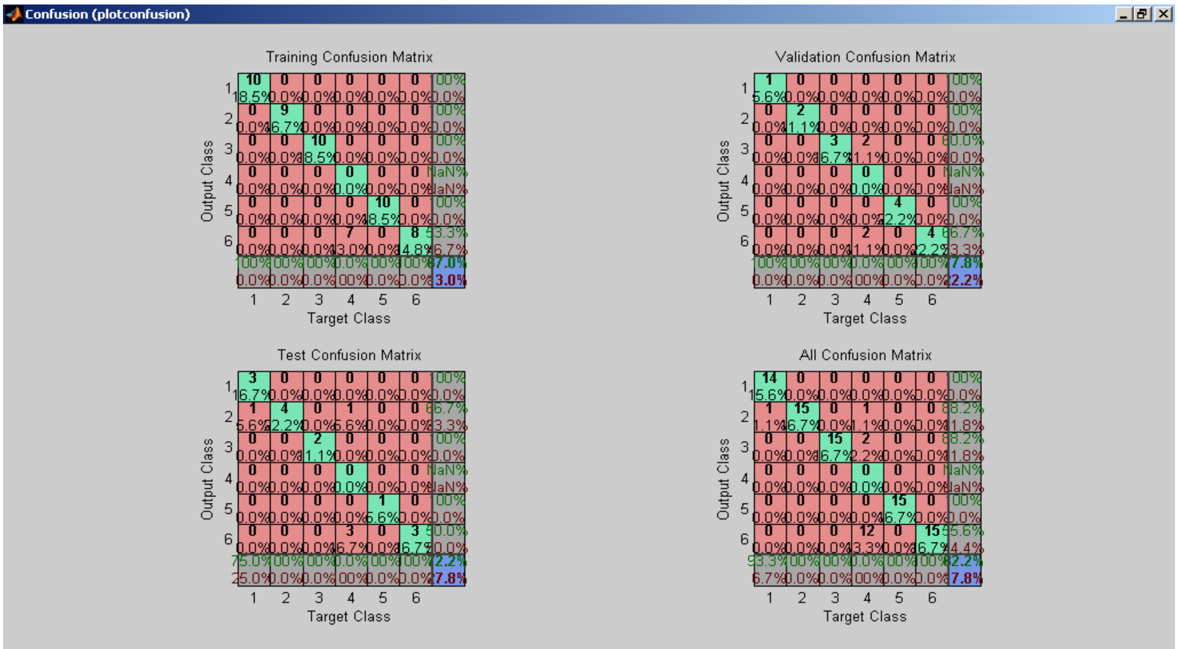

Fig.9 Confusion matrix for six class motor imagery moment

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VI. CONCLUSIONS

This paper presents the multiclass classification of six motor imagery classes including elbow, feet and wrist moment using DWT as feature extractor and ANN classifier. The overall accuracy comes out to be 82%. SVD method helps in achieving better accuracy.

ACKNOWLEGEMENT

This work is supported by Dr. B.R Ambedkar National Institute of Technology, Jalandhar. The authors would like to thank NITJ staff for providing the RMS 32 software for the successful recording of data and compilation of whole experiment.

REFERENCES

1. Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control Clin. Neurophysiol.
2. Vaughan T M et al 2003 Brain–computer interface technology: a review of the second international meeting IEEE Trans. Neural Syst. Rehabil. Eng.
3. H. H. Ehrsson, S. Geyer, and E. Naito, "Imagery of Voluntary Movement of Fingers, Toes, and Tongue Activate Corresponding Body-Part-Specific Motor Representations," *J Neurophysiol,* vol. 90, pp. 3304-3316, November 1, 2003 2003.
4. Baoguo Xu, Aiguo Song ,Juan Wu, "Algorithm of Imagined Left-right Hand Movement Classification Based on Wavelet Transform and AR -Parameter Model," , Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on 6-8 July 2007, pp. 539–542, .
5. G. Gage, K. Ludwig, K. Otto, E. Ionides, and D. Kipke. Naive coadaptive cortical control. J. Neural Eng., vol. 2, 2005, 52–63..
6. Mousavi, E.A.; Maller, J.J.; Fitzgerald, P.B.; Lithgow, B.J. Wavelet common spatial pattern in asynchronous offline brain computer interfaces. *Biomed. Signal Process. Control* **2011**, *6*, 121–128
7. Jiang, W.; Guizhi, X.; Lei, W.; Huiyuan, Z. Feature Extraction of Brain-Computer Interface Based on Improved Multivariate Adaptive Autoregressive Models. In Proceedings of the 3rd International Conference on Biomedical Engineering and Informatics (BMEI'10), Yantai, China, October 2010; pp. 895–898.
8. T. Wang, J. Denga, B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns", in Clinical Neurophysiology, Vol. 115, 2004, pp.2744–2753.
9. "The Wavelet Tutorial by Robi Polkar" http://users.rowan.edu/~polikar/WAVELETS/WTpar14.html
10. B. Scholkopf, A.J. Smola, Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press. London, England, 2002.
11. Blankertz B, Curio G and M¨uller K R 2002 Classifying single trial EEG: towards brain computer interfacing Adv. Neural Inf. Process. Syst. (NIPS 01) **14** 157–64
12. Fukunaga K 1990 Statistical Pattern Recognition 2nd edn (New York: Academic)
13. S. O. Haykin. Neural Networks and Learning Machines. Prentice Hall, 2008.
14. Ientilucci, E.J., (2003). \Using the Singular Value Decomposition". http://www.cis.rit.edu/~ejipci/research.htm
15. L. Qin and B. He. A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications, J. Neural Eng., vol. 2, no. 4, 2005, 65–72.
16. Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., and Kotchoubey, B. (2009). brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients.Ann. NY Acad. Sci. 1157, 90–100.
17. Krauledat M, Tangermann M, Blankertz B, Müller K-R, Towards Zero Training for Brain Computer Interfacing. 2008PLoS ONE 3(8): e2967.
18. S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller. Spatio-spectral filters for improved classification of single trial EEG.. IEEE Trans. Biomed. Eng., vol. 52, no. 9, Sep. 2005, 1541–1548.
19. L. Qin, L. Ding, and B. He. Motor imagery classification by means of source analysis for brain computer interface applications. J. Neural Eng., vol. 1, 2004, 135–141.
20. Lemm S, Schäfer C, Curio G, "BCI Competition 2003--Data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements," IEEE Trans. Biomed Eng., vol. 51, no. 6pp. 1077-1080, June 2004.

AUTHORS PROFILE

Rajesh Singla was born in Punjab, India in 1975. He obtained B.E Degree from Thapar University in 1997, M.Tech degree from IIT -Roorkee in 2006. Currently he is pursuing Ph.D degree from National Institute of Technology Jalandhar, Punjab, India. His area of interest is Brain Computer Interface, Rehabilitation Engineering, and Process Control. He is working as an Associate Professor in National Institute of Technology Jalandhar, India since 1998.

Neha Sharma was born in Jalandhar, Punjab, India. She received the B.E. in Instrumentation and Control Engineering from Dr. B.R Ambedkar National Institute of Technology, Jalandhar and pursuing M.Tech. degree in Control and Instrumentation Engineering from Dr. B.R Ambedkar National Institute of Technology, Jalandhar , India. Her research interests are in the areas of Biomedical Signal Processing.