Motor Imagery EEG signal Classification on DWT and Crosscorrelated signal features

Nischal K. Verma,Senior Member, IEEE, L.S.Vishnu Sai Rao and Suresh K. Sharma
Department of Electrical Engineering
Indian Institute of Technology Kanpur-208016, India

***Abstract*—Motor imagery (MI) based electroencephalogram (EEG) signals are a widely used form of input in brain computer interface systems (BCIs). Although there are a number of ways to classify data, a question still persists as to which technique should be employed in the domain of MI based EEG signals. In this paper, an attempt is made to find the best classification algorithm and feature extraction technique by comparing some of the prominently used algorithms on a same base dataset. Feature extraction techniques like discrete wavelet transform (DWT) and cross-correlation have been studied and compared. Five classification algorithms have been implemented which are logistic regression (LR), kernalised logistic regression (KLR), multilayer perceptron neural network (MLP), probabilistic neural network (PNN) and Least-square support vector machine (LS-SVM). Dataset IVa of BCI competition III has been used as a base dataset to test the algorithms. Evaluation of the algorithms has been done using a 10-fold cross-validation procedure. Experimental results show that a combination of DWT and LSSVM classifier outperforms the other procedures.**

Keywords: Brain Computer Interface (BCI), EEG signal, classification algorithms, feature extraction, cross-validation.

I. Introduction

Brain Computer Interface (BCI) Systems give venue for human brain to communicate with an external device via a non-physiological path. It replaces the use of nerves and muscles of the human brain with a combination of hardware and software, which translates electrophysiological signals into physical actions [1]. Motor imagery based BCI is a very productive communication method for people with motor disabilities. Motor Imagery (MI) is a mental process wherein the subject imagines that he is performing a specific motor action such as a hand or foot movement without otherwise performing it in reality [2]. Electroencephalogram (EEG) signals are used as inputs to BCI systems [3]. EEG signals are feature extracted in order to overcome the contaminations of noise and artifacts in them [4]. Machine learning algorithms [5] are then used in the classification of different brain patterns obtained upon performing different motor imagery tasks. Fig.1 shows a schematic diagram of signal processing done to an EEG recording, starting from signal acquisition through to classification outcomes.

Feature extraction and classification are the two main processes that are required to understand the EEG signals. There are a number of algorithms which could be used to implement these two processes. This paper is aimed to identify a good combination of feature extraction and classification technique among some of the most commonly used techniques. In this paper, feature extraction has been done using cross-correlation and discrete wavelet transform. The techniques used for classification are logistic regression (LR), kernalised logistic regression (KLR), probabilistic neural network (PNN), multilayer perceptron neural network (MLP) and Least-square support vector machine (LS-SVM). All these algorithms have been implemented on dataset IVa of BCI competition III. The algorithms have been assessed using a 10 fold cross validation method. Research done in this paper has been inspired by the work done by Siuly Siuly [6].

The rest of the paper is organized as follows. Section II outlines brief theory and implementation procedures of different algorithms used in this study. Section III gives the experiments which have been performed on the base dataset. Sections IV and V provide the observations and conclusions which can be drawn based on the experiments in Section III.

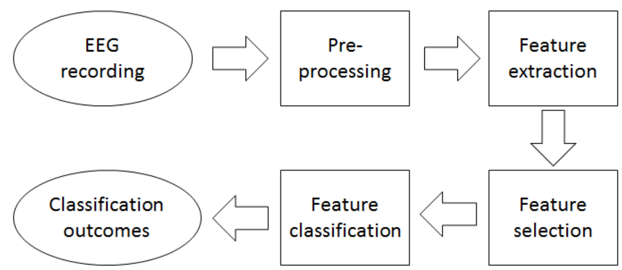

Fig. 1. Flow chart showing different stages of EEG signal processing

II. Theory and Implementation

A. Data Selection and Preprocessing

Dataset IVa of BCI competition III has been chosen for experimentation [7]. The dataset was comprised of five sub datasets. Each of the five datasets corresponded to the data acquired from five healthy subjects (aa, al, av, aw, ay). The subjects performed two classes of MI tasks which were right hand movement and right foot movement. All the five Datasets constituted of continuous signals from 118 EEG channels and so they were put in the form of a matrix of 118 columns. Data processing was required in order to classify the data systematically into the two classes. Thus a matrix of 118 columns was converted to a matrix of 236 columns of which the first 118 columns belonged to the first class and the last 118 columns belonged to the second class. Each column of

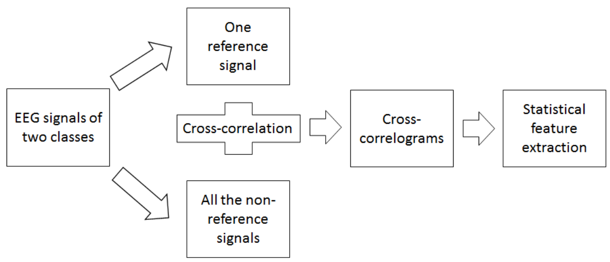

Fig. 2. Overview of Cross-correlation

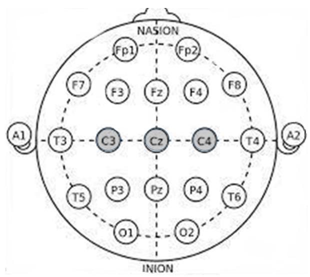

Fig. 3. MI Electrode position in 10/20 EEG electrode placement system

this matrix would represent a separate training example and the class to which the training example would belong to was also associated with it.

B. Feature Extraction

1) Cross-correlation Approach: Cross-correlation is a feature extraction technique used to find relationship (a sense of similarity) between two signals [8] .Cross-correlated signals are called cross-correlograms. A cross-correlogram is not only noise free but also contains greater amount of discriminative information which was otherwise absent in the original EEG signal [9].

A flowchart overviewing the process of cross-correlation is shown in Fig.2. The first step in cross-correlation is the selection of reference signal. EEG positions C3 and Fp1 in the international 10/20 system for electrode placement [10] are the two choices for a reference signal considered in this study. Fig3 shows the spatial position of some of the electrodes. Those which are darkened depict the most informative regions corresponding to MI data. After choosing the reference signal, cross-correlation is done between the reference signal and all the other non-reference signals, according to the equation

$$R_{xy}[m] = \sum_{i=0}^{N-|m|-1} x[i]y[i-m] \tag{1}$$

$$m = -(N-1), -(N-2), ...0, 1, 2, 3, ...(N-2), (N-1)$$

Hence if the first column i.e., the first training example, is cross-correlated with all the other 235 training examples, it would result in 235 new signals [11].

In order to handle the cross-correlogram sequence comfortably, the dimension of the data has to be reduced. Dimension of the data is reduced by extracting some statistical attributes from the cross-correlogram sequence. Features like maximum, minimum, mean, mode, median, standard deviation, $1/4$ percentile (first quartile), inter quartile range and $3/4$ percentile (third quartile) have been employed in this study. All these statistical attributes were evaluated using standard MATLAB commands except for mode where a histogram was first computed for the signal and then its peak was chosen as mode.

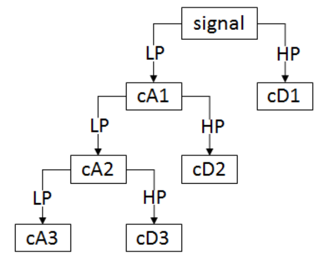

Fig. 4. DWT tree with 3 decomposition levels

2) Discrete Wavelet Transform: Discrete wavelet transform (DWT) is a feature extraction technique in which signal is decomposed using low pass and high pass filters [12]. The signal is decomposed to yield low frequency contents (approximation coefficients) and high frequency contents (detail coefficient) from a low pass filter and a high pass filter respectively. The approximations are passed again through low pass filter and high pass filter for decomposition into the next level. Fig.4 shows a diagrammatic version of DWT wherein the signal is decomposed to cA1, the approximate coefficient and cD1, the detailed coefficient. It is also observed that the tree is one sided i.e., approximate coefficients are further decomposed while the details are not. Several families of filters such as Daubechies, Coifflets, Haar can be used for decomposition. Among these Daubechies-4 (db4) decomposition filter has been considered in this study [13]. The other parameter to be set in DWT is the decomposition level which depends on the task and the signal to be performed. In this study, signals are decomposed to six levels resulting in six detail coefficients and one approximation coefficient. Nine features are to be extracted from these coefficients. The first three are variances of the detail coefficients at level 1, 2 and 3. Detail coefficients at level 4, 5 and 6 have been auto correlated. Variance of them was then calculated to give the next three features. The last three features were found by taking the absolute mean of smoothened versions of detail coefficients at level 1, 2 and 3 [14].

C. Feature Scaling

Feature scaling is a process of standardising the range of independent variables or data features. Z-transform has been used for feature scaling in this paper. In this technique, each feature is subtracted by its mean and divided by its standard deviation to give a new scaled version of the original feature set [15].

D. Evaluation

The performance of the classification algorithms is evaluated using 10 fold cross validation technique [16]. In this technique, the dataset containing several training examples is randomly split into ten mutually exclusive subsets of approximately same size. One of the 10 subsets is used for testing and all the others are put together for training. This process is repeated 10 times, using a different subset for testing every time. The ratio of number of times the prediction is violated to the total number testing samples in a particular trial gives the error for that trial. Thus for 10 trials, 10 errors are obtained whose mean is considered for measuring the overall accuracy of the algorithm.

E. Logistic Regression

Logistic regression (LR) [17] is a probabilistic statistical classification procedure used for predicting a dependent variable based on one or more independent variables. Suppose there are n observations $(y_1, x_1), (y_2, x_2), ...(y_n, x_n)$. Here $x_i \in \mathbb{R}^d$ are independent variables and y_i are Bernoulli distribution based binary variable with parameter p_i such that $p_i = Probability(y_i = 1)$. The binary responses can be modelled using the following formulation

$$\log \frac{p_i}{1-p_i} = \beta^T \mathbf{x}_i \qquad \text{or} \qquad p_i = \frac{\exp(\beta^T \mathbf{x}_i)}{1+\exp(\beta^T \mathbf{x}_i)} \tag{2}$$

where $\beta \in \mathbb{R}^d$ are unknown regression coefficients ordinarily estimated using maximum likelihood estimation [18]. Training and testing in this paper have been done using statistics toolbox in Matlab.

F. Kernal Logistic Regression

Kernel logistic Regression (KLR) [19] can be considered as a nonlinear version of Logistic Regression. There is a limitation in LR to classify the data with nonlinear boundaries. This problem can be solved using the kernel trick in KLR [20]. This method involves the use of a Kernel function. A possible choice to which is RBF (radial bias function) $K(x_i, x) = exp(-(||x_i - x||)^2/2\sigma^2)$. The kernels sensitivity can be controlled using the parameter σ. The parameters of this method can be estimated by Newton-Raphson method.
KLR has been implemented in this paper, using the 'minFunc' software provided by Mark Schmidt [21]. RBF was used as the kernel and the RBF scale was set as 1. After evaluating the kernel function, the parameters of the model were evaluated using the 'minFunc'. For testing the data, parameters of the model remain the same, change is only in the kernel function which is re-evaluated. The new kernel function along with the training parameters are used to predict the output class of the testing data.

G. Multilayer Perceptron Neural Network

Multilayer perceptron (MLP) is a feed forward neural network with an input layer, one or more hidden layers and an output layer [22]. Apart from the input layer, each neuron receives signals from the neurons of the previous layer weighted by the interconnected values between neurons. A neuron could be modelled according to a mathematical equation as follows

$$y = f\left(\sum_{i=1}^{n} w_{ji} x_i + b_i\right) \tag{3}$$

where $x_1, x_2, ...x_n$ are the inputs to the neuron, y is its output, w_{ij} is the weight , b_i is the bias term and $f(.)$ is a non-linear activation function (usually hyperbolic tangent sigmoid transfer function (tansig)) [23].A back propagation (BP) algorithm is used to train the MLP Neural Network. There are many different types of back propagation algorithm of which the Levenberg-Marquardt BP algorithm (LMA) is considered the quickest and most efficient and is used for training MLP in this study [24].
MLP has been implemented in this study using the Neural Network toolbox present in MATLAB. Training ratio was set as hundred percent because the training and testing dataset were already pre-determined for cross-validation. The number of hidden layers could be varied and different number of hidden layer yielded different accuracies. The number of hidden layers have been chosen so as to maximise the accuracy. Fig.5 shows a plot between variation of accuracy with the number of hidden layers. Subject 'al' was cross-correlated choosing C3 as a reference signal for making the plot. It can be observed from the plot that optimum number of hidden layers for this particular case is 15.

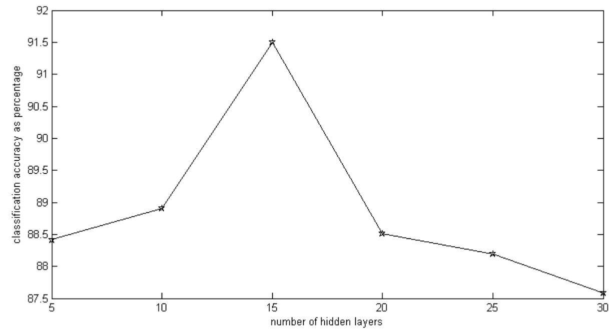

Fig. 5. Variation of accuracy with the number of hidden layers

H. Probabilistic Neural Network

Probabilistic neural network (PNN) has a feed forward architecture with an input layer followed by a hidden layer and an output layer. For classification, the data is presented at the input layer. Hidden layer evaluates the distance between the training input vectors and the input data [25] [26]. The computation of distance is done using a radial bias function ($K(x_i, x) = exp(-(||x_i - x||)^2/2\sigma^2)$).The contributions obtained from each training input vector are summed up at the output layer. Finally, the maximum of all the contributions is picked and thus the data is mapped to the corresponding class [27].
Training and testing have been done using the Neural Network toolbox present in MATLAB. RBF was used as a kernel

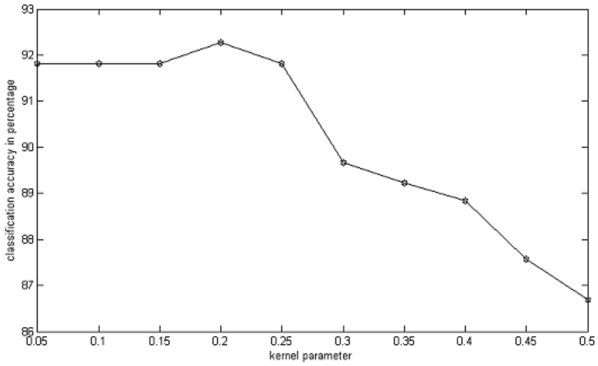

Fig. 6. Variation of accuracy with the RBF kernel parameter (σ)

function. Different kernel parameters resulted in different accuracies, hence a range from 0.05 to 0.5 was considered and that value was chosen which resulted in maximum accuracy. Fig. 6 shows a plot between variations of classification accuracy with the kernel parameter. Subject 'al' was cross-correlated, choosing C3 as a reference signal, for making the plot. From the plot it is observed that the kernel parameter value of 0.2 results in maximum accuracy for this particular case.

I. Least Square Support Vector Machine

Support Vector machine (SVM) is a most widely used machine learning algorithm for classification problems. An SVM creates a non-linear decision boundary by non-linearly mapping the input vector to a high dimensional feature space using a kernel function. A linear large margin classifier is constructed in the obtained feature space. The optimisation problem is often achieved by solving a quadratic programming (QP) problem [28] [29].

In the least square version of support vector machine, SVM problem is formulated by using a least square cost function [30] [6]. The difference between LS-SVM and SVM comes here while solving the optimisation problem. The optimisation problem can be solved in LS-SVM directly by solving a linear set of equations in contrast to the QP solving in SVM. Implementation of this algorithm was done using the toolbox LS-SVMlab version1.8 [31]. The kernel function used for LS-SVM in this study is radial bias function ($K(x_i, x) = exp(-(||x_i - x||)^2/2\sigma^2)$). LS-SVM is controlled by adjusting two parameters which are γ, which is the regularisation parameter, and σ^2, which is a parameter which controls the kernel's sensitivity. These parameters are determined in this study using a two-step grid search technique [32].The value of γ and σ^2 have to be low so as to avoid over fitting. The evaluated parameters γ and σ^2 were used in training in order to find the model parameters. A 10-fold cross-validation was done to measure the classification accuracy.

III. CASE STUDY

The experiments have been performed using MATLAB R2013a (Version 8.1). Three techniques are used for feature extraction, the first one is cross-correlation, discrete wavelet transform is the second one and the third one is a combination of cross-correlation and DWT. Five classification algorithms have been tested for each technique of feature extraction using 10-fold cross validation procedure.

1) Case 1: TableI shows the classification accuracies of the algorithms using cross-correlation as a feature extraction technique. Fp1 has been used as a reference signal and six statistical vectors have been extracted from the cross-correlograms which are mode, mean, standard deviation, minimum, median and maximum.

TABLE I
CROSS-CORRELATION BASED FEATURE EXTRACTION WITH FP1 AS REFERENCE SIGNAL AND 6 VECTOR FEATURE SET

	LR	KLR	MLP	PNN	LSSVM
aa	99.58%	100.0%	100.0%	99.57%	100.0%
al	87.68%	95.38%	96.23%	88.06%	99.17%
av	97.48%	99.13%	99.15%	99.55%	99.58%
aw	95.74%	93.17%	96.19%	89.80%	96.55%
ay	82.06%	82.15%	85.12%	82.19%	88.11%
avg	92.5	94.0	95.3	91.8	96.7

2) Case 2: TableII shows the classification accuracies of the algorithms using cross-correlation as a feature extraction technique. The same six features as in TableI are extracted from the cross-correlogram sequence but over here C3 is used as a reference signal.

TABLE II
CROSS-CORRELATION BASED FEATURE EXTRACTION WITH C3 AS REFERENCE SIGNAL AND 6 VECTOR FEATURE SET

	LR	KLR	MLP	PNN	LSSVM
aa	99.15%	99.58%	99.17%	99.58%	99.58%
al	92.35%	92.71%	91.50%	92.28%	94.81%
av	96.61%	97.43%	97.84%	97.43%	98.26%
aw	99.58%	96.14%	99.15%	94.47%	98.73%
ay	74.03%	85.52%	86.82%	80.86%	88.96%
avg	92.3	94.3	94.9	92.9	96.1

3) Case 3: TableIII shows the classification accuracies of the algorithms using cross-correlation as a feature extraction technique. Nine statistical features are extracted from the cross-correlograms. The first six features are same as that of TableI and TableII, the additional three are $1/4$ percentile (first quartile), inter quartile range and $3/4$ percentile (third quartile). Fp1 is used as a reference signal here.

4) Case 4: The classification accuracies of TableIV obtained using cross-correlation as feature extraction

TABLE III
CROSS-CORRELATION BASED FEATURE EXTRACTION WITH Fp1 AS REFERENCE SIGNAL AND 9 VECTOR FEATURE SET

	LR	KLR	MLP	PNN	LSSVM
aa	99.58%	100.0%	100.0%	99.57%	100.0%
al	86.35%	95.74%	93.26%	89.36%	99.58%
av	98.33%	99.13%	98.73%	99.55%	99.13%
aw	95.30%	91.88%	93.62%	89.74%	96.55%
ay	86.41%	87.68%	88.87%	82.26%	92.30%
avg	93.2	94.9	94.9	92.1	97.5

technique. The same features is used in TableIII. Reference signal chosen over here is C3.

TABLE IV
CROSS-CORRELATION BASED FEATURE EXTRACTION WITH C3 AS REFERENCE SIGNAL AND 9 VECTOR FEATURE SET

	LR	KLR	MLP	PNN	LSSVM
aa	99.57%	99.58%	99.58%	99.58%	99.58%
al	95.34%	93.13%	94.93%	89.21%	98.24%
av	96.14%	97.43%	97.84%	98.69%	97.84%
aw	98.73%	96.59%	99.58%	95.32%	97.84%
ay	74.88%	87.64%	89.78%	84.76%	91.94%
avg	92.9	94.9	96.3	93.5	97.1

5) Case 5: TableV shows the classification accuracies using Discrete wavelet transform as a feature extraction technique. Nine features are extracted here which are described in this secII-B2.

TABLE V
DISCRETE WAVELET TRANSFORM BASED FEATURE EXTRACTION

	LR	KLR	MLP	PNN	LSSVM
aa	99.58%	100.0%	99.58%	100.0%	99.58%
al	100.0%	99.17%	100.0%	99.58%	99.58%
av	98.33%	99.17%	99.17%	99.17%	99.17%
aw	100.0%	100.0%	100.0%	100.0%	100.0%
ay	97.86%	96.23%	98.32%	92.84%	98.75%
avg	99.2	98.9	99.4	98.3	99.4

6) Case 6: The classification accuracies shown in TabelVI are obtained by using cross-correlation along with discrete wavelet transform as feature extraction technique. The cross-correlograms obtained after cross-correlation are discrete wavelet transformed and 9 feature vectors have been extracted from the final sequence. The vectors which are extracted are describe in this Section II-B2. The reference signal used for cross-correlation is Fp1.

TABLE VI
FEATURE EXTRACTION BY CROSS CORRELATION AND DISCRETE WAVELET TRANSFORM WITH Fp1 AS REFERENCE SIGNAL

	LR	KLR	MLP	PNN	LSSVM
aa	100.0%	99.15%	100.0%	99.57%	99.57%
al	100.0%	99.13%	100.0%	99.17%	100.0%
av	92.79%	97.03%	94.89%	95.29%	97.88%
aw	97.84%	97.84%	97.90%	97.88 %	99.11%
ay	98.71%	99.13%	99.15%	99.58%	99.55%
avg	97.9	98.5	98.4	98.3	99.2

7) Case 7: TableVII is similar to TableVI except for the reference signal used for cross-correlation, which over here is C3.

TABLE VII
FEATURE EXTRACTION BY CROSS CORRELATION AND DISCRETE WAVELET TRANSFORM WITH C3 AS REFERENCE SIGNAL

	LR	KLR	MLP	PNN	LSSVM
aa	100.0%	99.17%	100.0%	99.58%	99.17%
al	100.0%	99.13%	100.0%	99.58%	100.0%
av	95.31%	94.94%	94.44%	95.76%	96.56%
aw	98.71%	98.67%	98.75%	97.82%	99.55%
ay	98.69%	99.13%	98.73%	99.17%	99.58%
avg	98.5	98.2	98.4	98.4	99.0

IV. OBSERVATIONS

A. Feature Extraction Technique

In order to compare the feature extraction techniques, the classification algorithm has to be the same. Average classification accuracies over all the tables have to be compared to find the best feature extraction procedure. The accuracies using LR algorithm from TableI to TableVII are 92.5%, 92.3%, 93.2%, 92.9%, 99.2%, 97.9% and 92.5% respectively. Similarly the accuracies of KLR are 94.0%, 94.3%, 94.9%, 94.9%, 98.9%, 98.5% and 98.2%. The accuracies of MLP algorithm from TableI to TableVII are 95.3%, 94.9%,94.9%, 96.3%, 99.4%, 98.4% and 98.4%. PNN algorithm gives accuracies as follows 91.8%, 92.9%, 92.1%, 93.5%, 98.3%, 98.3% and 98.4% from TableI to TableVII respectively. Finally the average accuracies using LSSVM algorithm are 95.5%, 96.1%, 97.5%, 97.1%, 99.4%, 99.2% and 99.0% from TableI to TableVII respectively. It is observed that DWT performs distinctively better for all classification algorithms except for PNN. In PNN there is a slight dominance of the technique used in TableVII over DWT. Thus it is reasonable to conclude that DWT can be a promising technique for feature extraction of MI based EEG signals.

B. Classification Algorithm

To compare the classification algorithms, the feature extraction technique has to be the same. Considering the average accuracies from TableII it is observed that LR gives 92.3%, KLR gives 94.3%, MLP gives 94.9%, PNN gives 92.9% and finally LSSVM gives an accuracy of 96.1%. It is observed from this data that LSSVM algorithm has given very fine results when compared to all the other algorithms. Similarly it is observed from all the tables that LSSVM gives better results among other algorithms irrespective of feature extraction techniques.

C. Number of Feature Vectors in Cross-correlation

Table pairs I and III, II and IV differ only in the number of time domain feature vectors extracted. For tables III and IV the accuracy, upon increasing the feature vectors from six to nine, has been improved for all the algorithms. For tables I and III it is the same situation except for the case of MLP. Thus it can be concluded that increasing the feature vectors has a slight improvement in accuracy.

V. Conclusion

In this paper, a comparative study of five classification methods and three feature extraction techniques of MI based EEG signal for BCI systems is provided. The classification performances have been assessed using a 10-fold cross validation technique under common ground conditions. Among the feature extraction techniques, it is DWT and among classification algorithms, it is LSSVM which are giving better classification results for EEG signal classification. Cross-correlation is giving good results but may not be preferred as it may cause the EEG signal to lose its temporal information. Overall, a combination of DWT and LSSVM seems to be a good procedure for classifying two class MI based EEG signals.

References

[1] "Braincomputer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767 – 791, 2002.

[2] R. Kottaimalai, M. Rajasekaran, V. Selvam, and B. Kannapiran, "Eeg signal classification using principal component analysis with neural network in brain computer interface applications," in *Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013 International Conference on*, March 2013, pp. 227–231.

[3] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust eeg single-trial analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41–56, 2008.

[4] S.-L. Wu, C.-W. Wu, N. Pal, C.-Y. Chen, S.-A. Chen, and C.-T. Lin, "Common spatial pattern and linear discriminant analysis for motor imagery classification," in *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2013 IEEE Symposium on*, April 2013, pp. 146–151.

[5] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial eeg during motor imagery by iterative spatio-spectral patterns learning (isspl)," *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 6, pp. 1733–1743, June 2008.

[6] S. Siuly and Y. Li, "Improving the separability of motor imagery eeg signals using a cross correlation-based least square support vector machine for brain," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 20, no. 4, pp. 526–538, July 2012.

[7] http://bbci.de/competition/iii/#references.

[8] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of {EEG} signals," *Expert Systems with Applications*, vol. 36, no. 2, Part 1, pp. 1329 – 1336, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417407005696

[9] G. M. Hieftje, R. I. Bystroff, and R. Lim, "Application of correlation analysis for signal-to-noise enhancement in flame spectrometry. use of correlation in determination of rhodium by atomic fluorescence," *Analytical Chemistry*, vol. 45, no. 2, pp. 253–258, 1973.

[10] H. H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.

[11] B. Hermanto, T. Mengko, A. Indrayanto, and A. Prihatmanto, "Brain signal reference concept using cross correlation based for brain computer interface," in *Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 2013 3rd International Conference on*, Nov 2013, pp. 388–391.

[12] C. Smith and N. Nair, "Comparing discrete wavelet transform (dwt) with discrete fourier transform (dft) implemented for digital relays," in *Power Engineering Conference, 2009. AUPEC 2009. Australasian Universities*, Sept 2009, pp. 1–6.

[13] S. Goumas, M. E. Zervakis, and G. S. Stavrakakis, "Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction." *IEEE T. Instrumentation and Measurement*, vol. 51, no. 3, pp. 497–508, 2002.

[14] N. Verma, S. Singh, J. Gupta, R. Sevakula, S. Dixit, and A. Salour, "Smartphone application for fault recognition," in *Sensing Technology (ICST), 2012 Sixth International Conference on*, Dec 2012, pp. 1–6.

[15] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, pp. 563–582, 2001.

[16] A. Sengur, "Multiclass least-squares support vector machines for analog modulation classification," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6681–6685, Apr. 2009.

[17] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.

[18] A. Subasi, A. Alkan, E. Koklukaya, and M. K. Kiymik, "Wavelet neural network classification of eeg signals by using ar model with mle preprocessing," *Neural Netw.*, vol. 18, no. 7, pp. 985–997, Sep. 2005.

[19] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[20] S. P. Rahayu, S. Purnami, and A. Embong, "Applying kernel logistic regression in data mining to classify credit risk," in *Information Technology, 2008. ITSim 2008. International Symposium on*, vol. 2, Aug 2008, pp. 1–6.

[21] http://www.di.ens.fr/~mschmidt/Software/minFunc.html.

[22] M. Riedmiller, "Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptive learning algorithms," 1994.

[23] N. Harun, S. Dlay, and W. Woo, "Performance of keystroke biometrics authentication system using multilayer perceptron neural network (mlp nn)," in *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, July 2010, pp. 711–714.

[24] M. Hagan and M.-B. Menhaj, "Training feedforward networks with the marquardt algorithm," *Neural Networks, IEEE Transactions on*, vol. 5, no. 6, pp. 989–993, Nov 1994.

[25] D. F. Specht, "Probabilistic neural networks," *Neural Netw.*, vol. 3, no. 1, pp. 109–118, Jan. 1990. [Online]. Available: http://dx.doi.org/10.1016/0893-6080(90)90049-Q

[26] P. Burrascano, "Learning vector quantization for the probabilistic neural network," *Neural Networks, IEEE Transactions on*, vol. 2, no. 4, pp. 458–461, Jul 1991.

[27] R. Lippmann, "Pattern classification using neural networks," *Communications Magazine, IEEE*, vol. 27, no. 11, pp. 47–50, Nov 1989.

[28] S. Agrawal, N. Verma, P. Tamrakar, and P. Sircar, "Content based color image classification using svm," in *Information Technology: New Generations (ITNG), 2011 Eighth International Conference on*, April 2011, pp. 1090–1094.

[29] N. Kohli, N. Verma, and A. Roy, "Svm based methods for arrhythmia classification in ecg," in *Computer and Communication Technology (ICCCT), 2010 International Conference on*, Sept 2010, pp. 486–490.

[30] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.

[31] http://www.esat.kuleuven.be/sista/lssvmlab/.

[32] M. Adankon, M. Cheriet, and A. Biem, "Semisupervised least squares support vector machine," *Neural Networks, IEEE Transactions on*, vol. 20, no. 12, pp. 1858–1870, Dec 2009.