Proceedings of the 2009 IEEE
International Conference on Information and Automation
June 22 -25, 2009, Zhuhai/Macau, China

Comparison of Different Classification Methods for EEG-Based Brain Computer Interfaces: A Case Study

Boyu Wang, Chi Man Wong, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai

Abstract—**The performances of different off-line methods for two different Electroencephalograph (EEG) signal classification tasks – motor imagery and finger movement, are investigated in this paper. The classifiers based on linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), kernel fisher discriminant (KFD), support vector machine (SVM), multilayer perceptron (MLP), learning vector quantization (LVQ) neural network, k-nearest neighbor (*k*-NN), and decision tree (DT), are compared in terms of classification accuracy. The main purpose of this paper is to provide a fair and extensive comparison of some commonly employed classification methods under the same conditions so that the assessment of different classifiers will be more convictive. As a result, a guideline for choosing appropriate algorithms for EEG classification tasks is provided.**

I. Introduction

Brain computer interface (BCI) is a communication system that allows its users to control external devices with brain activity, which does not depend on the brain normal output pathways of peripheral nerves and muscles [1], [2]. Currently, the electroencephalogram (EEG) signal, one of the non-invasive measurements of brain activity, is a most prevailing signal used in the BCI systems due to its excellent temporal resolution and usability. Therefore, BCI systems based on EEG are widely studied and a variety of algorithms have been proposed to indentify intended motions of the subjects in EEG recordings.

So far, some excellent surveys and comparison studies have been published. Müller and Anderson *et al* investigated the performances of linear and nonlinear methods in BCIs [3] [4], and linear methods were recommended, even though nonlinear methods can provide better results in some applications, especially with complex nonlinear EEG data set. Moreover, it is argued that one should always regularize to limit the complexity of the classifiers even when using linear methods – in particular for BCI data. This holds even more so for nonlinear methods [3] [5] [6]. Lotte *et al* [7] gave a comprehensive review of classification algorithms for EEG-based BCI. They identified the critical properties of different classifiers and provided guidelines to choose the suitable algorithms for a specific BCI system. Another thorough survey of signal processing algorithms in BCIs can be found in [8].

The evaluation of the performances of different EEG classification methods can be more convincing if they are measured under the same conditions. To date, some comparative studies of EEG classification methods for BCIs have been published [25] – [27]. To our best knowledge, however, no work has published to give a comprehensive comparison of the classifiers. This paper aims at filling this lack. The main purpose of this paper is to provide an objective and extensive comparison of some commonly employed classifiers and give a brief guideline to choose appropriate algorithms for EEG classification tasks. More specifically, the performances of the classifiers based on linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), kernel fisher discriminant (KFD), support vector machine (SVM), multilayer perceptron (MLP), learning vector quantization (LVQ) neural network, k-nearest neighbor (k-NN), and decision tree (DT) are assessed and compared. All of these methods are applied to two datasets – dataset III and IV of "BCI competition 2003". In addition, for these two dataset, all of the EEG raw data are processed by the same preprocessing and feature extraction methods respectively before the classification to guarantee the fair comparison.

The authors are with the Department of Electrical and Electronics Engineering, Faculty of Science and Technology, University of Macau, Av. Padre Tomás Pereira,Taipa, Macau. E-mail: fwan@umac.mo

II. Methods

A. Linear and Quadratic Discriminant Analysis

Given a set of observations in p-dimensional space: $D_i = \{\mathbf{x}_1^i, \dots, \mathbf{x}_{n_i}^i\}$ ($\mathbf{x}_j^i \in \mathbb{R}^p$) from class ω_i ($i = 1,\dots, C$, C is the number of classes), we assume that each of the class probability density functions can be modeled as normal distribution. Define the prior probabilities $p(\omega_i)$, means $\mathbf{m}_i$, and covariance matrices $\mathbf{\Sigma}_i$ of each class:

$$p(\omega_i) = \frac{n_i}{\sum_j n_j}$$
$$\mathbf{m}_i = \frac{1}{n_i}\sum_{i=1}^{n_i}\mathbf{x}_i \tag{1}$$
$$\mathbf{\Sigma}_i = \frac{1}{n_i}\sum_{i=1}^{n_i}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

where n_i is the number of patterns in class ω_i. Using Bayes rule and the normal assumption, we have the quadratic discriminant rule [9]: assign $\mathbf{x}$ to ω_i if $g_i > g_j$, for all $j \neq i$, where

$$g_i(\mathbf{x}) = \log(p(\omega_i)) - \frac{1}{2}\log(\mathbf{\Sigma}_i) - \frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\mathbf{m}_i) \tag{2}$$

Particularly, if we impose a stronger assumption that all the class covariance matrices $\mathbf{\Sigma}_i,...,\mathbf{\Sigma}_C$ are same, on (1), the discriminant function (2) simplifies and we will get the linear discriminant rule: assign $\mathbf{x}$ to ω_i if $g_i > g_j$, for all $j \neq i$, where g_i is the linear discriminant

$$g_i(\mathbf{x}) = \log(p(\omega_i)) - \frac{1}{2}\mathbf{m}_i^T\mathbf{S}_W^{-1}\mathbf{m}_i + \mathbf{x}^T\mathbf{S}_W^{-1}\mathbf{m}_i \tag{3}$$

in which $\mathbf{S}_W$ is the common covariance matrix

$$\mathbf{S}_W = \sum_{i=1}^{C}\frac{n_i}{n-C}\Sigma_i \tag{4}$$

For the special case of two classes, the linear discriminant function (2) can be written as: assign $\mathbf{x}$ to ω_1 if

$$\mathbf{w}^T\mathbf{x} + w_0 > 0 \tag{5}$$

else assign $\mathbf{x}$ to ω_2, where in the above

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{6}$$

$$w_0 = -\log\left(\frac{p(\omega_1)}{p(\omega_2)}\right) - \frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} \tag{7}$$

LDA and QDA have been applied successfully in a large number of BCI systems [3] [4] [6] [10] [11], mainly because that the assumptions of LDA and QDA have been fulfilled in some BCI experiments. But for complex nonlinear EEG data, they may provide poor result [12].

B. Kernel Fisher Analysis

With the help of Mercer kernels, one can express $\mathbf{w}$ in terms of mapped training patterns [13] [14]

$$\mathbf{w} = \sum_{i=1}^{n}\alpha_i\mathbf{\Phi}(\mathbf{x}_i) \tag{8}$$

and perform the dot product directly in high dimensional feature space by using the kernel trick. The projection of an input data onto the discriminant is compute by

$$(\mathbf{w}\cdot\mathbf{\Phi}(\mathbf{x})) = \sum_{i=1}^{n}\alpha_i K(\mathbf{x}_i,\mathbf{x}) \tag{9}$$

where $K(\mathbf{x}_i,\mathbf{x}) = (\mathbf{\Phi}(\mathbf{x}_i)\cdot\mathbf{\Phi}(\mathbf{x}))$ is a mercer kernel computing a scalar product in feature space. The discriminant function therefore can be written as

$$f(x) = \sum_{i=1}^{n}\alpha_i K(\mathbf{x}_i,\mathbf{x}) + w_0 \tag{10}$$

which results in nonlinear discriminant boundaries, and the bias of the classifier can be chosen as the mean of the average projections of two classes. In this study, Gaussian kernel is adopted as the kernel function.

C. Support Vector Machine

The aim of LDA/KFD is to maximize the separation between the projected class means while also minimize the variance within each class. Different from LDA/KFD, SVM selects the hyperplane with the best generalization capabilities by maximizing the margins, i.e., the distances from the nearest training data. This leads to different objective functions of LDA/KFD and SVM. The optimal hyperplane of SVM can be obtained by solving a constrained optimization problem [14]

$$\min_{\mathbf{w},w_0,\xi} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{11}$$

$$\text{subject to} \quad \xi_i \geq 0,\ y_i\left((\mathbf{w}\cdot\mathbf{\Phi}(\mathbf{x}_i)) + w_0\right) \geq 1 - \xi_i .$$

where ξ_i is the slack variables, $C > 0$ is the regularization constant determining the tradeoff between the empirical error and the complexity term. This results in the dual problems

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{\mathrm{n}}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j) \tag{12}$$

$$\text{subject to} \quad 0 \leq \alpha_i \leq C,\ i = 1,...,n,$$

$$\sum_{i=1}^{n}\alpha_i y_i = 0$$

where α_i is the Lagrange multiplier. The training samples for which Lagrangian multiplier is not zero are called support vectors.

By using different types of kernel functions, one can obtain different discriminant boundaries. If the scale product is adopted as the kernel function, the corresponding SVM creates linear discriminant boundaries and nonlinear otherwise. In this comparison study, scale product and the Gaussian kernel are adopted as the kernel functions, and the corresponding SVM are known as linear SVM and Gaussian SVM respectively, and for Gaussian SVM, the scaling factor in the kernel function is chosen as 100.

D. Neural Networks

Neural networks are composed of several neurons which are used to develop nonlinear classification boundaries. An MLP is a simple type of neural networks, which includes at least one hidden layer of neurons. The most widely employed training strategy for MLP is backpropagation, which uses gradient descent applied to a sum-of-squares error function, and the corresponding network is usually called as backpropagation network (BPN).

It should be noted that the term *BPN* is often used to describe multilayer perceptrons employing such a calculation, though strictly *backpropagation* refers to the method of derivative calculation, rather than the type of network [9].

Another category of neural networks mostly employed in BCI system design is the LVQ network. Different from the BPN, the LVQ network applies winner-take-all Hebbian learning-based algorithm. As its name indicates, it is related to vector quantization methods in which the input space is divided into a number of regions, where each region has a reference vector and a class label attached. In the classification stage, the input vector is classified to the class label of the reference vector which is closest to this input vector [15]. Learning in the LVQ network is essentially to find the reference vectors.

In this study, both MLP and LVQ are three layers neural networks, and the numbers of neurons in hidden layer are 8 and 6 respectively.

E. k-Nearest-Neighbor

This algorithm assigns an input data to the dominant class among its k nearest neighbors within the training set [16]. These neatest neighbors are usually determined by Euclidean distance of the input data. For the two data sets in this study, k is selected as 11.

The main drawback of this technique is its sensitivity to the curse-of-dimensionality, which may be usually encountered in the EEG classification task. However, by using appropriate feature extraction technique, we can apply k-NN successfully in the BCI system with low dimensional feature space, as being demonstrated in section IV.

F. Decision Tree

Decision tree is one of the most popular classification algorithms in Data Mining and Machine Learning because it is easily interpreted and comprehended by humans. The tree T is made up of nodes and branches. Node t is designated as an internal node or a terminal node (leaf node). Internal node can split into two children (t_L and t_R) while terminal node does not have any children. A terminal node has a class label associated with it such that observations that fall into particular terminal node are assigned to that class [16] – [19].

The DT is obtained by splitting nodes iteratively. The main problem is when to stop the splitting during the growing. To avoid overfitting, we first generate a fully grown tree, in which each terminal node only contains one training data; then, the meaningless braches from this large tree. As a result, a sequence of subtrees can be obtained by successively pruning branches, and the optimal tree is selected by cross-validation.

III. EXPERIMENTS

A. Data Collection

The performances of the classification methods are compared against two datasets: dataset III and IV in BCI competition 2003, which are provided by Graz BCI group and Fraunhofer-FIRST, Intelligent Data Analysis Group; Freie Universität Berlin, Neurophysics Group respectively.

Dataset III was recorded from a 25 years old female during a feedback session. Three pairs of bipolar EEG channels (anterior '+', posterior '-') were positioned over C3, Cz and C4. It was sampled at 128 Hz and bandpass filtered between 0.5 and 30Hz. This dataset consists of 280 trials (140 left trials and 140 right trials) of 9 s length. In each trial, the first 2s was quite. At 3s, an arrow (left or right) was displayed as cue while the female was asked to imagine left or right hand movement according to the cue. One half of dataset (140 trials: 70 left trials and 70 right trials) are used to training the BCI system, the others are used to testing.

Dataset IV was recorded from a normal subject during a no-feedback session. The task is to press with the index and little fingers the corresponding keys in a self-chosen order and timing 'self-paced key typing'. The EEG was collected by 28 electrodes at the positions of the international 10/20-system. The duration of the signal is 500ms ending 130 ms before a keypress, and the sample rate was 100Hz. There're 416 trials in the dataset, including 316 training trials and 100 testing trials.

B. Data Preprocessing and Feature Extraction

Two categories of feature extraction strategies are applied in this study due to the differences of the two EEG datasets.

1) Dataset III: For this task, we introduced event-related desynchronization (ERD) on the contralateral sensorimotor area during motor imagery.

ERD of *mu* and *beta* rhythms is a distinct characteristic for imagery movement classification task. One of the basic measurements of ERD/ERS in time domain is that the EEG power within identified frequency bands is displayed relative to the power of the same EEG derivations recorded during the reference or baseline period a few seconds before the event occurs introduced in [20]. This is a very convenient, fast and simply method to extract the ERD feature from filtered EEG data.

The dataset III was sampled from 3 channels where are C3, Cz and C4, respectively. Only C3 and C4 channel data are employed here since the ERD feature in Cz is not clear. First, the C3 and C4 channel data is filtered by IIR (Infinite Impulse Response) bandpass filter (10-12 Hz). Then the two time courses of ERD $e_{c3}(k)$ and $e_{c4}(k)$ are computed at k-th sampling point by equation (13) [21]. $x_{c3}(k)$ and $x_{c4}(k)$ are the filtered data from C3 and C4 channel, respectively,

$$e_{c3}(k) = \sum_{i=0}^{w-1} x_{c3}(k-i)^2 \ , \ e_{c4}(k) = \sum_{i=0}^{w-1} x_{c4}(k-i)^2 \tag{13}$$

where w is the length of time window. In this experiment, we choose w = 15 (≈ 117.2 ms), and starting point of the time window is t = 560 (= 4.375s). Each feature vector $e_c(k)$ is constructed by two column vectors $e_{c3}(k)$ and $e_{c4}(k)$, where is denoted in (14).

$$e_c(k,m) = [e_{c3}(k,m), e_{c4}(k,m)] \tag{14}$$

where k is denoted the k-th sampling point and m is denoted the m-th trial.

2) Dataset IV: For finger movement task, we extract the features from another category of electrophysiological activities, i.e. bereitschaftspotential (BP).

BP, also named readiness potential (RP), as a component of movement-related potentials (MRPs), is low-frequency potential that reflects the dynamic changes in motor cortical activity prior to the movement onset. They have bilateral distribution and present maximum amplitude at vertex. Close to the movement, they become contralaterally preponderant [2]. Thus, the feature extracted from BP can be utilized in the finger movement task.

Since the BP of finger movement dominates in the low frequency band, a low-pass filter is applied to extract the BP of the finger movement from the EEG. The cutoff frequency is 7Hz, which was adopted in the previous work [23]. It should be noted that the filter used here is the zero-phase filter to avoid phase shift.

For dataset IV, the starting point and the length of the time window are t = 43 (200ms before the keypress) and w = 7 (60ms) respectively.

In order to utilize more information, we make use of all the channels of dataset IV, and therefore common spatial subspace decomposition (CSSD) [22], one category of spatial filters, is employed to combine all the electrodes to process multi-channel EEG. The aim of CSSD is to separate the evoked responses and background spontaneous brain activities (specific and common activities), which are overlapped in the scalp measurement. Given single-trial multichannel spatial-temporal EEG signal matrices X_L and X_R (evoked by left and right finger movements respectively) with dimension N (channels) by T (samples), they can be modeled as follows:

$$X_L = [C_L \ C_C]\begin{bmatrix} S_L \\ S_C \end{bmatrix} \quad X_R = [C_R \ C_C]\begin{bmatrix} S_R \\ S_C \end{bmatrix} \tag{15}$$

where C_L and C_R are the spatial patterns related to left and right finger movements respectively, and C_C represents the spatial pattern specific to the background activities. Then S_L, S_R, and S_C are the corresponding source activities related to the left and right figure movements, and the common condition. Our aim is to construct spatial filters F_L and F_R by using CSSD to extract source activities:

$$S_L = F_L X \quad S_R = F_R X \tag{16}$$

Then for each testing data X, we define $f = [x_L \ x_R]$ as the feature vector, where $x_L = F_L \cdot X$ and $x_R = F_R \cdot X$. It has been shown [23] that x_L of left trials and x_R of right trials have larger amplitudes than that of the contrary patterns, which makes the high accuracy classification possible; for more details see [22] – [24].

C. Model Selection

Model selection plays a vital role in the performance of a classifier. In our experiment, the optimal parameters of the classifiers mentioned previously are selected by 4-fold cross-validation on the training data. The cross-validation is performed 12 times, and the average classification accuracy is adopted as the criterion for choosing the optimal parameters of the classifiers.

IV. Results and Discussions

The performances of aforementioned classifiers in terms of classification accuracy are shown in table I. Datasets I and II are datasets III and IV in BCI competition 2003 respectively.

From table I, it can be observed that the Gaussian SVM and k-NN achieve desirable performance in both datasets, while LVQ and QDA show the lowest accuracies for dataset I and dataset II respectively. Additionally, KFD and MLP also have inferior results.

As linear classifiers, LDA and linear SVM have similar performances. Since the linear classifiers are generally more robust than their nonlinear counterparts due to their limited flexibility and insensitivity to overfitting, the use of linear methods is recommended whenever possible [3].

Generally, k-NN classifier is not very popular in the BCI community, probably because they are known to be very sensitive to the curse-of-dimensionality. However, as demonstrated in table I, if appropriate feature extraction methods are applied, the dimension of the input vector will be much lower, so the k-NN classifier can achieve the good performance. In our experiment, after the feature extraction, the dimension of the input vector for dataset I is two, and fourteen for dataset II, which have been reduced significantly, and the k-NN classifier therefore is efficient.

It should be noted that the classification accuracies of DT are also satisfactory, especially for dataset II. Additionally, DT can be easily extended for multiclass pattern recognition task. Unfortunately, to our knowledge, few publications have applied this classifier to BCI research.

TABLE I
CLASSIFICATION ACCURACY OF DIFFERENT CLASSIFICATION METHODS

Classification Method	Classification Accuracy	
	Dataset I (Motor Imagery)	Dataset II (Finger Movement)
LDA	82.86%	84%
QDA	78.57%	79%
KFD	80.71%	81%
Linear SVM	82.86%	82%
Gaussian SVM	84.29%	84%
MLP	80.71%	81%
LVQ	77.86%	80%
k-NN	84.29%	83%
DT	82.14%	86%

As mentioned above, the accuracies of the classifiers are computed based on fixed time window. In general, the parameters of the temporal filter should be considered carefully. They can be selected by relevant prior knowledge or some other techniques e.g. cross validation. Table II and table III demonstrate the classification accuracies of the classifiers with different starting points and lengths of the time window. The best performance of each classifier is shown in bold font. Overall, the optimal starting point for dataset I is between 550-560 and, 43-46 for dataset II.

From these two tables, we can have some further understanding of the properties of the classifiers and the effects of the parameters of the time window. SVMs (both linear and Gaussian) achieve desirable and robust classification accuracies in both datasets with different time windows. This probably is due to their immunity to the curse-of-dimensionality and regularization property [7]. However, LVQ and QDA seem to be more sensitive to different parameters of temporal filter.

It should be noted that each classifier achieves its own highest accuracy with different time windows, which means when designing a BCI system, one should consider the feature extraction and classification algorithm together rather than independently.

V. Conclusion

In this paper, we provide the comparisons between nine classification methods in EEG-based BCI systems. The performances of the classifiers are assessed with the same datasets, data preprocessing and feature extraction methods.

Linear classifiers are generally the first choice for EEG signal classification due to their simplicity, stability and insensitivity to overfitting. If regularization is applied, the

TABLE II
CLASSIFICATION ACCURACY WITH DIFFERENT PARAMETERS OF TEMPORAL FILTER (DATASET I)

Classification Method	Starting Point and Length of Time Window (t, w)								
	(560,16)	(607,29)	(550,39)	(557,22)	(563,29)	(547,32)	(551,7)	(553,7)	(554,31)
LDA	**85.00%**	82.14%	82.86%	82.14%	82.14%	82.86%	81.43%	83.57%	82.86%
QDA	78.57%	**83.57%**	76.43%	77.86%	77.14%	75.51%	77.86%	77.86%	77.14%
KFD	81.43%	80.71%	**84.29%**	80.00%	80.00%	82.14%	80.00%	80.00%	80.00%
Linear SVM	83.57%	81.43%	82.86%	**85.00%**	83.57%	82.86%	82.14%	81.43%	81.43%
Gaussian SVM	84.29%	81.43%	82.86%	82.86%	**85.71%**	82.86%	81.43%	83.57%	83.57%
MLP	79.29%	80.00%	83.57%	81.43%	81.43%	**85.71%**	81.43%	84.29%	79.29%
LVQ	73.57%	71.43%	75.71%	78.57%	75.71%	70.71%	**83.57%**	72.86%	70.71%
k-NN	80.00%	78.57%	82.14%	83.57%	84.29%	83.57%	82.86%	**87.14%**	82.86%
DT	84.29%	78.57%	81.43%	81.43%	80.71%	78.57%	77.14%	82.86%	**86.43%**

TABLE III
CLASSIFICATION ACCURACY WITH DIFFERENT PARAMETERS OF TEMPORAL FILTER (DATASET II)

Classification Method	Starting Point and Length of Time Window (t, w)								
	(42,8)	(46,5)	(45,5)	(44,5)	(43,5)	(43,4)	(45,6)	(44,7)	(43,7)
LDA	**87%**	79%	82%	81%	82%	80%	79%	83%	84%
QDA	79%	**83%**	79%	82%	77%	76%	78%	82%	80%
KFD	78%	81%	**85%**	84%	82%	83%	81%	83%	81%
Linear SVM	83%	81%	82%	**86%**	84%	83%	82%	82%	82%
Gaussian SVM	85%	82%	83%	85%	**87%**	84%	83%	82%	83%
MLP	78%	80%	77%	81%	81%	**83%**	76%	78%	81%
LVQ	75%	81%	80%	77%	75%	73%	**84%**	82%	83%
k-NN	80%	83%	82%	81%	83%	82%	84%	**86%**	82%
DT	80%	81%	83%	81%	82%	82%	81%	82%	**86%**

nonlinear classifiers can achieve satisfactory result, otherwise they may be sensitive to noise and outliers, and therefore have poor performances. When proper feature extraction methods are employed to reduce the dimension of the input vector, *k*-NN is also an efficient algorithm in BCI research. In BCI community, few publications have applied DT to BCI system. But the results of our experiments have demonstrated that DT is also a promising technique for BCI system.

The effects of the parameters of temporal filter (the starting point and the length of the time window) are also investigated in this paper. Overall, when designing a BCI system, one should consider the feature extraction and the classification algorithms together rather than independently.

ACKNOWLEDGMENT

The authors gratefully acknowledge support from the Macau Science and Technology Department Fund and the University of Macau Research Fund (Grant RG059/08-09S/FW/FST).

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interface for communication and control," *Clin. Neurophsiol.,* vol. 133, pp. 767-791, June 2002.

[2] B. Blankertz *et al*., "The BCI competition III: validation alternative approach to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, pp. 153-159, June 2006.

[3] K.-R.Müller, C.W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 165–169, June 2003.

[4] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol.11, pp. 141-144, June 2003.

[5] K.-R.Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz, "Machine learning techniques for brain-computer interfaces," *Biomed. Technol., vol. 49, pp.11-22, 2004.*

[6] B. Blankertz, G. Curio, and K.-R. Müller, "Classifying single trial EEG: towards brain computer interfacing," in *Advances in Neural Information Processing Systems*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, 2002, vol. 14.

[7] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, pp. R1–R13, 2007.

[8] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, pp. R32–R57, 2007.

[9] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. New York: Wiley, 2002.

[10] G.A. Barreto, R.A. Frota, and F.N.S.de Medeiros, "On the classification of mental tasks: a performance comparison of neural and statistical approaches," in *Proc. 14th IEEE Workshop on Machine Learning for Single Processing*, 2004, pp. 529-538.

[11] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings," *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 1209-1214, Dec. 1990.

[12] G. N. Garcia, T. Ebrahimi, and J.-M. Vesin, "Support vector EEG classification in the Fourier and time-frequency correlation domains," in *Proc. 1st Int. IEEE EMBS Conf. Neural Engineering*, 2003, pp. 591-594.

[13] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Proc. IEEE Neural Networks Signal Processing Workshop*, 1999, pp. 41–48.

[14] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, pp. 181-201, Mar. 2001.

[15] J. Kalcher, D. Flotzinger, and G. Pfurtscheller, "A new approach to a brain-computer interface (BCI) based on learning vector quantization (LVQ3)," in *Proc. Annu. Int. IEEE EMBS Conf.* 1992, pp. 1658-1659.

[16] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001

[17] L. Breiman, J.H. Friedman, R.A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

[18] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.,* vol. 21, pp. 660-674, May-June 1991.

[19] I. H. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Mateo, CA: Morgan Kaufmann, 2005.

[20] G. Pfurtscheller, and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol*., vol. 110, pp. 1842–1857, Nov. 1999.

[21] W. Jia, X. Zhao, H. Liu, X. Gao, S. Gao, and F. Yang, "Classification of Single Trial EEG during Motor Imagery based on ERD," in *Proc. 26th Annu. Int. IEEE EMBS Conf.* 2004, pp. 5-8.

[22] Y. Wang, P. Berg, and M. Scherg, ''Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions: a simulation study," *Clin. Neurophysiol.,* vol. 110, pp. 604-614, Apr. 1999.

[23] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao, and F. Yang, "BCI competition 2003 data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1081-1086, June 2004.

[24] Y. Li, X. Gao, H. Liu, and S. Gao, "Classification of single-trial electroencephalogram during finger movement," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1019-1025, June 2004.

[25] F. Cincotti *et al*., "Comparison of different feature classifiers for brain computer interfaces," in *Proc. 1st Int. IEEE EMBS Conf. Neural Engineering*, 2003, pp. 645-647.

[26] H. Mirghasemi, R. Fazel-Rezai, M. B. Shamsollahi, "Analysis of P300 classifiers in brain computer interface speller," in *Proc. Annu. Int. IEEE EMBS Conf.* 2006, pp. 6205-6208.

[27] N. S. Dias, M. Kamrunnahar, P. M. Mendes, S. J. Schiff, and J. H. Correia, "Comparison of EEG pattern classification methods for brain-computer interfaces," in *Proc. Annu. Int. IEEE EMBS Conf.* 2007, pp. 2540-2543.