International Journal of Neural Systems, Vol. 0, No. 0 (XX, 2013) 00–00

AGGREGATION OF SPARSE LINEAR DISCRIMINANT ANALYSES FOR EVENT-RELATED POTENTIAL CLASSIFICATION IN BRAIN-COMPUTER INTERFACE

YU ZHANG
Key Laboratory for Advanced Control and Optimization for Chemical Processes
East China University of Science and Technology, Shanghai, China
E-mail: zhangyu0112@gmail.com

GUOXU ZHOU
Laboratory for Advanced Brain Signal Processing
RIKEN Brain Science Institute, Wako-shi, Japan
E-mail: zhouguoxu@brain.riken.jp

JING JIN
Key Laboratory for Advanced Control and Optimization for Chemical Processes
East China University of Science and Technology, Shanghai, China
E-mail: jinjingat@gmail.com

QIBIN ZHAO
Laboratory for Advanced Brain Signal Processing
RIKEN Brain Science Institute, Wako-shi, Japan
E-mail: qbzhao@brain.riken.jp

XINGYU WANG
Key Laboratory for Advanced Control and Optimization for Chemical Processes
East China University of Science and Technology, Shanghai, China
E-mail: xywang@ecust.edu.cn

ANDRZEJ CICHOCKI
Laboratory for Advanced Brain Signal Processing
RIKEN Brain Science Institute, Wako-shi, Japan
and
Systems Research Institute
Polish Academy of Science, Warsaw, Poland
E-mail: a.cichocki@riken.jp

Received (to be inserted
Revised by Publisher)

Two main issues for event-related potential (ERP) classification in brain-computer interface application are curse-of-dimensionality and bias-variance tradeoff, which may deteriorate classification performance, especially with insufficient training samples resulted from limited calibration time. This study introduces an aggregation of sparse linear discriminant analyses (ASLDA) to overcome these problems. In the ASLDA, multiple sparse discriminant vectors are learned from differently l_1-regularized least-squares regressions by exploiting the equivalence between LDA and least-squares regression, and are subsequently aggregated to form an ensemble classifier, which could not only implement automatic feature selection for dimensionality reduction to alleviate curse-of-dimensionality, but also decrease the variance to improve generalization capacity for new test samples. Extensive investigation and comparison are carried out among the ASLDA, the ordinary LDA and other competing ERP classification algorithms, based on different three ERP datasets. Experimental results indicate that the ASLDA yields better overall performance for single-trial ERP classification when insufficient training samples are available. This suggests the proposed ASLDA is promising for ERP classification in small sample size scenario to improve the practicability of brain-computer interface.

Keywords: Aggregation; Brain-computer interface (BCI); Electroencephalogram (EEG); Event-related potential (ERP); Sparse linear discriminant analysis.

1. Introduction

Brain-computer interface (BCI) is a technology to establish direct connection between a human brain and a computer without muscle activity, and hence provide a new communication channel for those people who have severe motor disabilities [1−5]. In recent years, BCI has been developed mainly based on three types of brain activities measured using electroencephalogram (EEG), i.e., event-related potential (ERP) [6,7], sensorimotor rhythm [8,9], and steady-state visual evoked potential [10,11]. Since the ERP-based BCI usually gives relatively more robust performance for target classification, it has aroused more interest of researchers [12,13].

ERPs are brain activities time- and phase-locked to experimental events of interest and contain typically N200, P300, etc. ERP-based BCI can detect desired commands from the subject through classifying the stimuli that elicit the largest amplitudes of ERPs. P300 has been most widely employed for the ERP-based BCI development, which is a positive deflection in EEG over parietal cortex, occurring approximately 300 ms after a rare but task-relevant stimulus presentation in an oddball paradigm. More recently, several variant ERP-based BCIs have been designed by exploiting N200, N170 and vertex positive potential associated with higher level cognitive function, namely motion perception and face perception [14,15].

ERPs are relatively weak and usually occur amid the ongoing background activities in the brain [16−20]. How to design an effective algorithm for ERP classification is one of the key issues for the development of improved ERP-based BCIs. Two main problems likely to deteriorate the ERP classification in BCI application are so-called curse-of-dimensionality and bias-variance tradeoff [21]. The curse-of-dimensionality is caused by the fact that samples (i.e., feature vectors) for ERP classification are generally extracted by concatenation of multiple time points from multiple channels and have typically high dimensionality, whereas the number of training samples is relatively small within limited calibration time [22,23]. This will most probably result in a poorly trained classifier giving high misclassification rate. On the other hand, the bias-variance tradeoff restricts generalization capability of the classifier. Especially, a limited training set will likely cause overfitting to outliers in EEG signals and hence a more severe bias-variance tradeoff. The classifier with poor generalization capability could hardly give good classification results on test data.

Until now, various algorithms have been proposed for EEG analysis, such as wavelet transform, neural network, fuzzy logic, support vector machine, and linear discriminant analysis (LDA), and so on [24−38]. Since LDA generally provides good results with low computational requirement, it has been more widely adopted [36−38]. However, LDA may get overfitting to high dimensional and possibly noised data, especially in small training sample size [39].

In recent years, regularization technique has shown its strength on preventing overtraining [41,42], and the LDA with regularization is probably the most promising approach to improve the ERP classification performance [39,43]. Typically, two regularized versions of the LDA, namely stepwise LDA (SWLDA) and Bayesian LDA (BLDA) have been most used for ERP classification in the BCI application [12,39,40]. The SWLDA is commonly employed to alleviate effects of small sample size on LDA transforms, and is less likely to corrupt the classification accuracy using fewer training samples since those insignificant features are removed from the discriminant model by forward and backward stepwise analysis with statistical tests [43]. As another regularized version of the LDA, the BLDA [39] prevents overfitting to possible outliers through Bayesian linear regression with hyperparameters estimated from the evidence framework [44]. More recently, a regularized LDA with shrinkage technique, called shrinkage LDA (SKLDA), was proposed to mitigate effects resulting from the high-dimensionality of features compared to the number of training samples on EEG classification [45]. The SKLDA remedies the ill-conditioned covariance matrix through the shrinkage covariance estimator [46], and effectively enhances generalization capability of classifier, thereby giving good ERP classification performance even when using insufficient training samples [7].

In this study, we propose an aggregation of sparse linear discriminant analyses (ASLDA) to ERP classification in small sample size scenario for BCI application. First, a sparse LDA (SLDA) is introduced by l_1-regularized least squares regression (LSR) based on the equivalence between LDA and LSR, to learn a sparse discriminant vector with capability of automatic feature selection for dimensionality reduction, and hence curse-of-dimensionality mitigation. Such l_1-regularized least squares method has been previously discussed in other BCI works [47−49]. The regularization parameter in the SLDA can be typically estimated by cross-validation scheme that however may cause overfitting to small training sample size due to the aforementioned bias-variance tradeoff, since there will be insufficient training samples for construction of training and test subset to select the regularization parameter effectively [50]. Combination of multiple classifiers has been suggested for overcoming the bias-variance tradeoff [21,41]. Thus, we utilize the conception of ensemble learning and propose further an aggregation scheme for regularization parameter selection in the SLDA. The ASLDA is subsequently obtained, in which a number of differently l_1-regularized LSRs are aggregated to form an ensemble classifier that could effectively reduce the variance to improve the generalization performance on test samples. The ASLDA is extensively validated on three ERP datasets: BCI Competition III-dataset II, EPFL's dataset and dataset recorded from our own experiment, and compared to the ordinary LDA, SLDA with cross-validation for regularization parameter selection (CVSLDA), SWLDA, BLDA and SKLDA. Experimental results show that the proposed ASLDA provides better overall performance when insufficient training samples are available, and demonstrates its potential for reducing calibration time of BCI system.

2. Methodology

2.1. *LDA and Relationship with LSR*

Linear discriminant analysis (LDA) is a benchmark method to find the optimal linear combination of features separating two classes [51,52]. Since the LDA has relatively low computational requirement and usually provides good classification results, it has been widely used for EEG classification [36−38]. Assume we are given a set of EEG training samples (i.e., feature vectors) $\mathbf{x}_i \in \mathbf{R}^D (D = MP)$ $(i = 1, 2, \ldots, N)$ where each sample is the concatenation of P temporal points from each of M channels, and the corresponding class labels $y_i \in \{1, -1\}$. The mean vectors of the two classes are computed as:

$$\mathbf{m}_1 = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \mathbf{x}_i, \ \mathbf{m}_2 = \frac{1}{N_2} \sum_{i \in \mathcal{I}_2} \mathbf{x}_i, \tag{1}$$

where $\mathcal{I}_1$ and N_1 are the index set and the number of samples in class 1 while $\mathcal{I}_2$ and N_2 are those of samples in class -1. The common covariance matrix is estimated as:

$$\mathbf{\Sigma} = \frac{1}{N} \sum_{c=1}^{2} \sum_{i \in \mathcal{I}_c} (\mathbf{x}_i - \mathbf{m}_c)(\mathbf{x}_i - \mathbf{m}_c)^T. \tag{2}$$

Then, the discriminant vector of the LDA is given by:

$$\tilde{\mathbf{w}} = \mathbf{\Sigma}^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \tag{3}$$

One potential problem for the LDA is that $\mathbf{\Sigma}^{-1}$ becomes ill-defined when the number of training samples is small compared to the dimensionality of feature space, and this will most probably give poor classification results [21,39]. An effective calibration for the classifier requires generally at least five to ten times as many training samples per class as the dimensionality, which is however difficult to meet in real application of the ERP-based BCI taking into account the system practicability [21,54].

LDA can be actually obtained from a special case of LSR for binary classification problem [51,53]. Assume the samples has been centered. LSR is to minimize the following cost function:

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{X}^T\mathbf{w} - \mathbf{y}\|_2^2, \tag{4}$$

where $\|\cdot\|_2$ denotes the l_2-norm, $\mathbf{X} \in \mathbf{R}^{D\times N}$ is a matrix obtained from the horizontal stacking of samples, i.e., $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$, and $\mathbf{y} \in \mathbf{R}^N$ denotes a vector containing the regression targets. Assume the regression targets in LSR have been specified to N/N_1 for samples from class 1 and to $-N/N_2$ for samples from class -1, respectively. By setting the derivative of Eq. (4) with respect to $\mathbf{w}$ to zero, we obtain:

$$\mathbf{X}\mathbf{X}^T\mathbf{w} = \mathbf{X}\mathbf{y}. \tag{5}$$

After some straightforward algebra, Eq. (5) follows:

$$(N\mathbf{\Sigma} + \frac{N_1N_2}{N}(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T)\mathbf{w} = N(\mathbf{m}_1 - \mathbf{m}_2). \tag{6}$$

Since $(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w}$ is always in the direction of $\mathbf{m}_1 - \mathbf{m}_2$, the optimal solution is then given by:

$$\tilde{\mathbf{w}} = Q\mathbf{\Sigma}^{-1}(\mathbf{m}_1 - \mathbf{m}_2), \tag{7}$$

where Q is a constant. That is, the LSR with regression targets as $y_i \in \{N/N_1, -N/N_2\}$ is equivalent to the LDA since irrelevant scale factor Q could be ignored for the discriminant vector.

LDA in the LSR framework allows us to impose various regularization operators more tractably on the discriminant model for different applications. Typically, l_2-regularization (Tikhonov regularization) [55], l_1-regularization (LASSO) [56], and l_1+l_2-regularization (elastic net) [57] can be used, in which the l_1-regularization is most commonly applied to learn sparse projection vector for dimensionality reduction and feature selection.

2.2. *Sparse LDA*

With the equivalent condition $y_i \in \{N/N_1, -N/N_2\}$ between LDA and LSR, a sparse linear discriminant analysis (SLDA) [47] is introduced through minimizing the following cost function:

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{X}^T\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_1, \tag{8}$$

where, $\|\cdot\|_1$ denotes the l_1-norm, λ is a positive regularization parameter to control the sparsity of discriminant vector $\mathbf{w}$, and larger λ would result in more sparse $\mathbf{w}$. Different from the standard LASSO problem, the specified targets $y_i \in \{N/N_1, -N/N_2\}$ maximize the discriminative information in regression. Effective feature selection for dimensionality reduction could be achieved with an appropriately chosen λ. To minimize the cost function in Eq. (8), we adopts the coordinate descent algorithm proposed by Friedman et al. [58], which provides a fast and accurate estimation for the sparse solution. Assume $\mathbf{X}$ is standardized so that $\sum_i x_{ji} = 0$, $\sum_i x_{ji}^2 = 1$. Suppose we have estimates $\tilde{w}_0$ and $\tilde{w}_l$ for $l \neq j$, and we wish to partially optimize $J(\mathbf{w})$ with respect to w_j. Then, $J(\mathbf{w})$ can be further written as:

$$J(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \tilde{y}_i^{(j)} - x_{ji}w_j)^2 + \lambda\sum_{l\neq j}|\tilde{w}_l| + \lambda|w_j|, \tag{9}$$

where $\tilde{y}_i^{(j)} = \sum_{l\neq j} x_{li}\tilde{w}_l$ is the fitted value excluding the contribution from x_{ji}, and $y_i - \tilde{y}_i^{(j)}$ is thus partial residual for fitting w_j. By the derivative of $J(\mathbf{w})$ with respect to w_j, the coordinate-wise update form is given by:

$$\tilde{w}_j \leftarrow S\left(\sum_{i=1}^{N} x_{ji}(y_i - \tilde{y}_i^{(j)}),\ \lambda\right), \tag{10}$$

where $S(a, \lambda)$ is a shrinkage-thresholding operator with value:

$$\operatorname{sign}(a)(|a| - \lambda)_+ = \begin{cases} a - \lambda & \text{if } a > \lambda \\ 0 & \text{if } |a| \leq \lambda \\ a + \lambda & \text{if } a < -\lambda. \end{cases} \tag{11}$$

Repeat the update form in Eq. (10) for $j = 1, 2, \dots, D, 1, 2, \dots$ until convergence, we then obtain the optimized discriminant vector $\tilde{\mathbf{w}}$ satisfying Eq. (8). The given λ determines sparsity degree of the $\tilde{\mathbf{w}}$, and hence the extent of feature selection for dimensionality reduction.

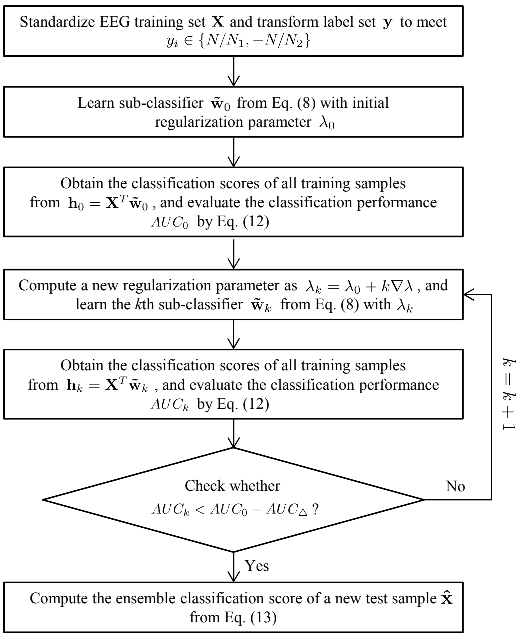

Fig. 1. Flowchart of the proposed ASLDA algorithm for ERP classification.

2.3. *Aggregation of SLDAs*

An appropriate λ can be selected by typically using cross-validation scheme, which gives the best classification performance on training set. However, the cross-validation may not be applicable in small training sample size, since there will be insufficient training samples for construction of training and test subset to effectively estimate λ [50]. Recently, ensemble learning methods have been suggested to be promising for improving performance of EEG feature extraction [50,59]. Accordingly, instead of the cross-validation, this study introduces an aggregation scheme based on the principle of ensemble learning to determine regularization parameter in the SLDA, and hence an aggregation of SLDAs (ASLDA) for ERP classification. In the proposed ASLDA, multiple sub-classifiers are first trained by Eq. (8) from a set of regularization parameters that are selected according to a pre-specified performance tolerance on the training set. Classification score of a new test sample is then obtained by the aggregation of sub-classification scores from all the trained sub-classifiers. Instead of a single regularization parameter, various regularization parameters result in different discriminant models. Such diversity is good for ensemble learning according to the generalization theory that interprets the success of boosting [60,61].

Fig. 1 shows flowchart of the ASLDA for ERP classification. Given EEG training set $\mathbf{X} \in \mathbf{R}^{D\times N}$, corresponding label set $\mathbf{y} \in \mathbf{R}^N$ where each label has been transformed to satisfy the equivalence between LDA and LSR, i.e., $y_i \in \{N/N_1, -N/N_2\}$, and an initial regularization parameter λ_0, we derive a sub-classifier $\tilde{\mathbf{w}}_0$ from Eq. (8) and compute the classification scores of all training samples from $\mathbf{h}_0 = \mathbf{X}^T\tilde{\mathbf{w}}_0$. The classification performance AUC_0 of all training samples is then evaluated by area under the receiver operating characteristic curve (AUC) [62]:

$$AUC = \frac{\sum_{i\in\mathcal{I}_1} r_i - \frac{N_1(N_1+1)}{2}}{N_1N_2}, \tag{12}$$

where $\mathcal{I}_1$ is the class with label N/N_1 and r_i is the rank of ith score when sorting all classification scores in ascending (i.e., rank of the minimal score is 1 and rank of the maximal score is N). Specifying an increasing step length $\nabla\lambda$ for aggregation and a performance tolerance $AUC_\triangle$, we repeat the following steps for $k = 1, 2, \ldots$, till $AUC_k < AUC_0 - AUC_\triangle$ is met:

- **Step 1** Compute a new regularization parameter as $\lambda_k = \lambda_0 + k\nabla\lambda$, and obtain the kth sub-classifier $\tilde{\mathbf{w}}_k$ from Eq. (8).
- **Step 2** Compute the classification scores of all training samples from $\mathbf{h}_k = \mathbf{X}^T\tilde{\mathbf{w}}_k$.
- **Step 3** Evaluate the classifier performance AUC_k of all training samples by Eq. (12).

Assume K sub-classifiers have been obtained from the aforementioned repeat, the classification score of a new test sample $\hat{\mathbf{x}} \in \mathbf{R}^D$ is then estimated from an aggregation solution based on all of the trained sub-classifiers:

$$\hat{h} = \sum_{i=0}^{K-1} \tilde{\mathbf{w}}_i^T\hat{\mathbf{x}}. \tag{13}$$

Here, the classification score of each test sample is simply obtained from the sum rule in Eq. (13) while more advanced combination scheme such as boosting [61] could also be adopted, and may improve the classification performance further. The initial regularization parameter λ_0 should be relatively small other than zero. The performance tolerance $AUC_\triangle$ controls the extent of aggregation. In this study, the parameters for aggregation scheme are empirically set as: $\lambda_0 = 0.005$, $\nabla\lambda = 0.005$ and $AUC_{\triangle} = 0.05$.

2.4. *Contrast Methods*

To validate effectiveness of the proposed ASLDA, it is compared with LDA, SLDA with cross-validation for regularization parameter selection (CVSLDA) and the other three competing ERP classification algorithms:

(a) Stepwise LDA (SWLDA). The SWLDA is commonly employed to alleviate effects of small sample size on the LDA transform through removing those insignificant features from the discriminant model by a combination of forward and backward stepwise analysis with statistical tests [43,63]. As three free parameters, two p-values controlling adding and removing of features, and the maximal number of features have to be predefined for the SWLDA. The parameter setting (add feature: p-value $<$ 0.1; remove feature: p-value $>$ 0.15; maximal number of features: 60) has been recommended by Krusienski et al. [63] for P300 classification. The same setting was adopted in this study.

(b) Bayesian LDA (BLDA). As an extension of the LDA, the BLDA can prevent overfitting to high dimensional and possibly noised datasets by regularization under Bayesian analysis [39]. The BLDA is to perform regression formulated by Eq. (4) in a Bayesian model where hyperparameters are estimated from the evidence framework [44].

(c) Shrinkage LDA (SKLDA). The SKLDA is a modified LDA by adjusting the extreme eigenvalues of the covariance matrix towards the average eigenvalue through the shrinkage covariance estimator [46], which can strengthen the generalization capacity of trained classifier for test data [7,45].

3. Experiments and Results

3.1. *EEG Data Acquisition*

3.1.1. *Dataset-1*

The Dataset-1 is from dataset II of the BCI competition III (http://www.bbci.de/competition/iii/) and provided by Wadsworth Center, Albany, NY. EEG signals were recorded at 240 Hz sampling rate from 64-scalp positions with high-pass and low-pass filters 0.1 Hz and 60 Hz. A 6×6 matrix consisting of characters was presented to the subject as the stimulus paradigm on a computer screen. The row/column of the matrix was intensified for 100 ms with an interstimulus interval (ISI) of 75 ms, and the subject was asked to focus attention on the cued character. Row and column intensifications were block randomized in blocks of 12. Only intensifications of the row and column containing the cued character should elicit ERPs, such as N200 and P300. The desired character can be determined through detecting the row and column with the largest ERP responses. The sets of intensifications were repeated 15 times and thus a total of 180 intensifications were presented to the subject for each character spelling. EEG data from two subjects (subject A and B) were available, and each subject had a train dataset containing 85 characters and a test dataset containing 100 characters. See [64] for more detailed description of this dataset.

Since this study focuses mainly on the problem of small sample size, we only use the train datasets of the two subjects for our analysis. The 16 channels F3, Fz, F4, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, PO7, PO8 and Oz are used for the subsequent feature extraction and classification. A 700 ms data segment is extracted from the beginning of each intensification and band-pass filtered from 1 Hz to 20 Hz by a sixth order forward-backward Butterworth bandpass filter. Each extracted segment is downsampled to 40 Hz by selecting each 6th point from the filtered data, thus the signal of each channel consists of 28 points. A total of 15300 samples (i.e., feature vectors) with dimensionality of $16 \times 28 = 448$ are extracted from each subject, and each 180 samples correspond to one character spelling.

3.1.2. *Dataset-2*

The Dataset-2 is from the EPFL BCI group (http://bci.epfl.ch/p300) and recorded by Hoffmann et al [39]. EEG data of eight subjects were available and recorded using a Biosemi Active Two amplifier (BioSemi, Netherland) at 2048 Hz sampling rate from 32 channels placed at the standard positions of the 10-20 international system. Each subject completed four experimental sessions and each session consisted of six runs. Stimuli consisted of six images and one of them was cued as the target stimulus in each run, where the six images started flashing randomly preceded by the target cue to elicit ERPs, such as N200 and P300. Each flash lasted for 100 ms with an ISI of 300 ms and subjects were asked to focus attention on the cued target stimulus. A total of 120 to 150 flashes were presented in each run.

EEG signals from the eight channels Fz, Cz, P7, P3, Pz, P4, P8 and Oz are used for analysis, which has been recommended by Hoffmann et al [39]. A 1000 ms data segment is extracted from each image flash, band-pass filtered from 1 Hz to 12 Hz by a sixth order forward-backward Butterworth bandpass filter, and downsampled from 2048 Hz to 32 Hz by selecting each 64th sample. Finally, 2880 to 3600 samples with dimensionality of $8 \times 32 = 256$ are obtained from each subject.

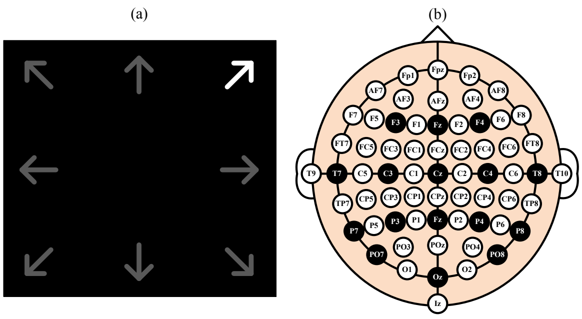

Fig. 2. Experimental paradigm (a) and channel configuration (b) for recording of the Dataset-3.

3.1.3. *Dataset-3*

The Dataset-3 was recorded from our own experiment. Ten healthy volunteers (S1-S10, aged from 22 to 49, one female) participated in the experiment. The subjects were seated in a comfortable chair 60 cm from a 17 inch LCD monitor (60 Hz refresh rate and 1280×1024 screen resolution) in a shield room. EEG signals were recorded at 256 Hz sampling rate using the g.USBamp amplifier (g.tec, Austria) with high-pass and low-pass filters of 0.1 Hz and 30 Hz. The following 16 electrodes (according to the 10-20 international system) were used for signal recording and analysis: F3, Fz, F4, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, PO7, PO8 and Oz. The experimental layout consisted of eight arrow stimuli to simulate a wheelchair navigation controller. Fig. 2 depicts the experimental paradigm and channel configuration. In the experiment, each subject completed two recording sessions, where each session consisted of eight runs. In each run, a randomly cued target arrow was first presented for 1 s in the middle of the screen followed by a 1 s black screen period, and a block-randomized sequence of stimulus intensifications was subsequently started. The intensification block was repeated five times, in each of which each arrow was intensified once with a duration of 100 ms and an ISI of 80 ms. During the experiment, subjects were asked to focus attention on the target arrows and silently count the number of times they were intensified so that ERPs, such as N200 and P300 were elicited.

A 700 ms data segment after baseline corrected by 100 ms pre-stimulus interval is extracted from each stimulus intensification and band-pass filtered from 1 Hz to 10 Hz by a sixth order forward-backward Butterworth bandpass filter. A total of 640 such segments consisting of 80 targets and 560 non-targets are derived for each subject from the two experimental sessions, and each 40 segments (5 targets and 35 non-targets) correspond to one command selection (16 command selections in total). Each segment is then downsampled to 21 Hz after 12-point moving average, thus the signal of each channel consists of 15 points. Finally, we obtain 640 samples with dimensionality of $16 \times 15 = 240$ from each subject.

3.2. ***Experimental Results***

In this study, we implemented an extensive investigation on the performances of LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA for single-trial ERP classification in small sample size scenario using the Dataset-1, Dataset-2 and Dataset-3. AUC computed according to Eq. (12) with test data was adopted to evaluated the ERP classification performance. Target detection accuracy with single trial was also evaluated from the Dataset-3. One-way analysis of variance (ANOVA) was used to evaluate performance difference among the classification algorithms.

3.2.1. *Classification Results for Dataset-1*

The samples extracted from spellings of 5 to 20 characters were randomly selected as for classifier training, while those from spellings of the remaining 65 characters for test to evaluate the classification performance. This procedure was repeatedly executed for 100 times and averaged AUC was then computed. Fig. 3 depicts the AUCs on test data derived by the LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA, respectively, when using various numbers of training characters (each character contains 180 samples) for classifier calibration. The proposed ASLDA performed better over the others when fewer training samples were available. Such superiority decreased gradually with the increasing of number of training samples.

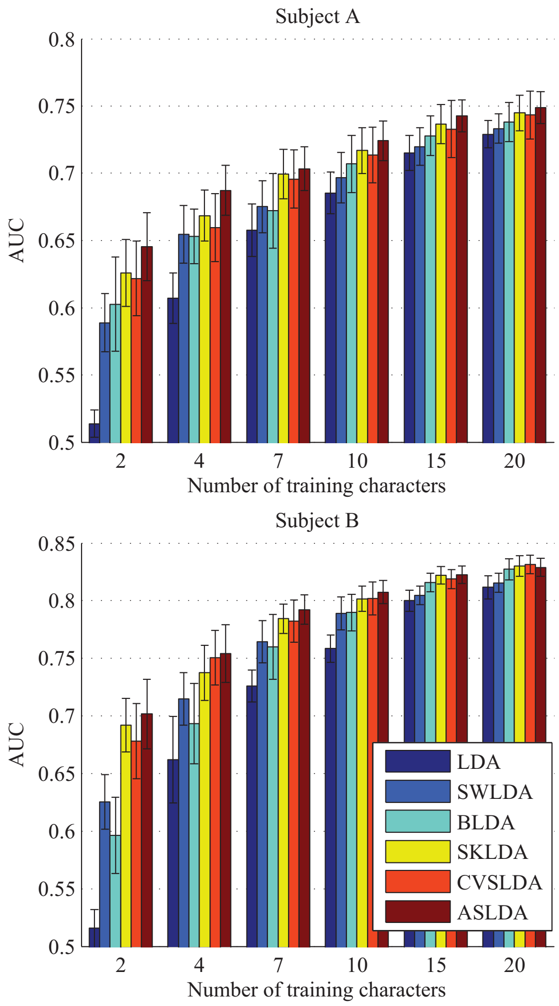

Fig. 3. Area under the ROC curve (AUC) on test data, derived by the LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA, respectively, using various numbers of training characters (each character contains 180 samples) for subjects A and B of the Dataset-1. The errorbar denotes the standard deviation of AUC on the 100 executions.

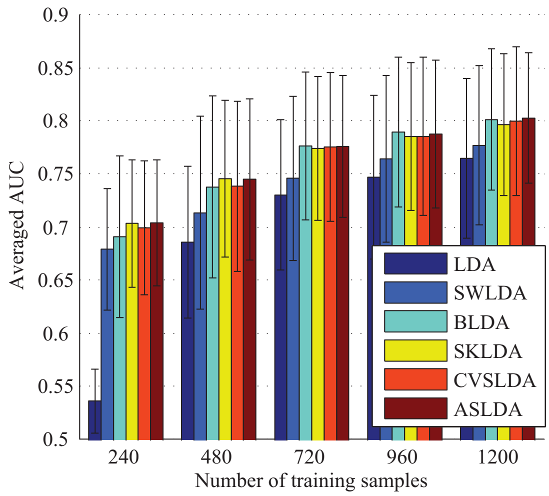

Fig. 4. Averaged area under the ROC curve (AUC) on test data, obtained by the LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA, respectively, using different numbers of training samples for the Dataset-2. The errorbar denotes the standard deviation of AUC on all subjects.

3.2.2. *Classification Results for Dataset-2*

The Dataset-2 was original published with proposition of the BLDA, which has proved superiority of the BLDA over the LDA for target detection in ERP-based BCI [39]. In this study, we analyzed performances of the LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA for single-trial ERP classification using different numbers of training samples, especially small training sample size. For each subject, 240 to 1200 samples were randomly selected for classifier training while the remaining samples for classification performance evaluation. This procedure was repeatedly executed for 100 times and averaged AUC was then computed. Fig. 4 shows the AUCs averaged on the eight subjects for test data, and obtained by the six algorithms, respectively, using 240 to 1200 training samples with an interval of 240. With 240 training samples, the ANOVA revealed a significant performance difference among the six algorithms ($F(5, 42) = 9.83$, $p < 0.0001$) . All of the modified versions of LDA achieved significantly better classification performance than that of the ordinary LDA (SWLDA $>$ LDA: $F(1, 14) = 28.7$, $p < 0.001$, BLDA $>$ LDA: $F(1, 14) = 39.1$, $p < 0.001$, SKLDA $>$ LDA: $F(1, 14) = 49.7$, $p < 0.0005$, CVSLDA $>$

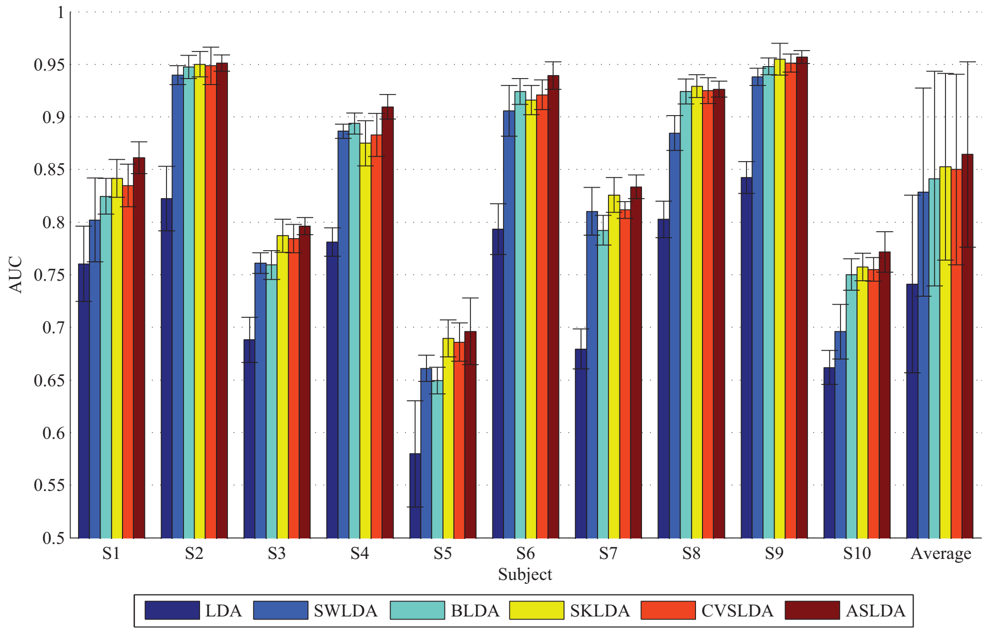

Fig. 5. Area under the ROC curve (AUC) on test data, derived by the LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA, respectively, and evaluated with 100 random selections of training samples from the Dataset-3. The errorbar of each subject denotes the standard deviation of AUC on the 100 executions. The errorbar of average result denotes the standard deviation on all subjects.

LDA: $F(1,14) = 43.7$, $p < 0.001$, ASLDA > LDA: $F(1,14) = 50.8$, $p < 0.0005$). No significant difference was found among the five modified LDA algorithms ($F(4,35) = 0.22$, $p = 0.93$).

3.2.3. *Classification Results for Dataset-3*

For the Dataset-3, half samples selected randomly from 640 samples were used for classifier training and the remaining half for test. This procedure was repeatedly executed for 100 times and averaged AUC was then computed. Hence, the number of training samples 320 (40 targets and 280 non-targets) was insufficient compared to the feature dimensionality of 240 (16 channels$\times$15 points). Fig. 5 depicts the AUCs derived by the six classification algorithms for the ten subjects, respectively. The ANOVA revealed a performance difference approaching significant among the six classification algorithms ($F(5,54) = 2.37, p = 0.051$). All of the five modified versions of LDA yielded significantly better classification performance than that of the ordinary LDA (SWLDA > LDA: $F(1,18) = 4.51$, $p < 0.05$, BLDA > LDA: $F(1,18) = 5.71$, $p < 0.05$, SKLDA > LDA: $F(1,18) = 8.29$, $p < 0.05$, CVSLDA > LDA: $F(1,18) = 7.75$, $p < 0.05$, ASLDA > LDA: $F(1,18) = 10.16$, $p < 0.01$). No significant performance difference is found among the five modified algorithms ($F(4,45) = 0.2$, $p = 0.94$). Table 1 shows the target detection accuracies obtained by the six classification algorithms with single trial for the Dataset-3. Although no significant difference was found among the six algorithms ($F(5,54) = 1.61$, $p = 0.17$), the ASLDA indeed achieved higher classification performance than the other five algorithms for most subjects.

In summary, the SWLDA, BLDA, SKLDA, CVSLDA and ASLDA significantly outperformed the ordinary LDA for single-trial ERP classification. Although no significant difference was found among the five modified LDA algorithms, the proposed ASLDA provided the best overall performance when insufficient training samples were available. The superior performance of the ASLDA over that of the CVSLDA indicated that the designed aggregation scheme is more effective than the cross-validation for regularization parameter selection of the SLDA in

Table 1. Target detection accuracies (%) obtained by the LDA, SWLDA, BLDA, SKLDA, CVSLDA and ASLDA with single trial for S1-S10 of the Dataset-3. For each subject, the best result is displayed in bold characters. The chance level is 12.5 %.

Subject	LDA	SWLDA	BLDA	SKLDA	CVSLDA	ASLDA
S1	38.5	46.3	48.5	53.3	51.5	**57.3**
S2	54.3	75.5	78.8	**80.5**	79.5	**80.5**
S3	33.8	37.5	36.5	44.5	43.3	**46.0**
S4	50.5	68.3	71.3	67.3	67.8	**73.8**
S5	18.8	23.3	22.5	28.8	26.5	**30.0**
S6	45.5	65.0	72.5	70.0	71.8	**77.5**
S7	29.5	44.3	40.0	48.8	44.5	**49.0**
S8	50.5	63.0	78.3	**79.8**	78.0	78.3
S9	56.0	79.3	81.3	**82.0**	80.0	81.8
S10	27.0	31.0	36.3	38.3	37.8	**40.5**
Average	40.4±12.8	53.4±19.4	56.6±22.0	59.3±19.1	58.1±19.6	**61.5**±19.2

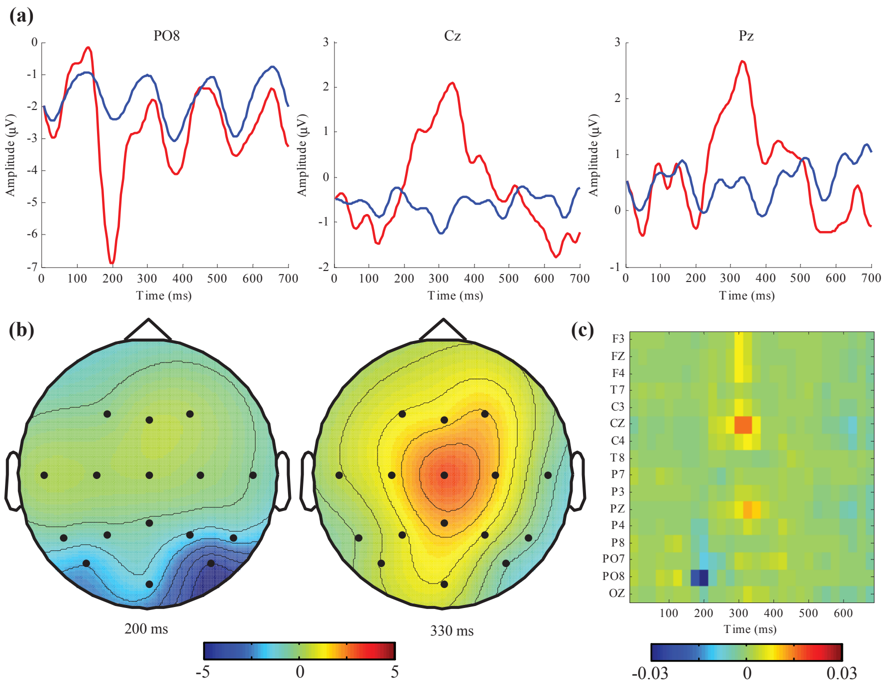

Fig. 6. ERP characteristics obtained from subject B of the Dataset-1 (using 10 training characters, i.e., 1800 samples). (a) Averaged ERP waveforms at electrodes PO8, Cz and Pz; (b) Scalp topographies of EEG amplitude differences between target and nontarget at 200 ms and 330 ms; (c) Discriminative information evaluated by signed r^2-values.

small sample size scenario.

4. Discussion

So far, a lot of algorithms have been introduced to EEG analysis for various applications, such as diagnosis of Alzheimer's disease, functional community of brain, and so on [65−75]. In the context of BCI, linear discriminant analysis (LDA) has been most popularly adopted for ERP classification since it usually provides good results and has relatively low computational requirement [63]. However, LDA is likely to give poor results due to the curse-of-dimensionality, since the number of training samples

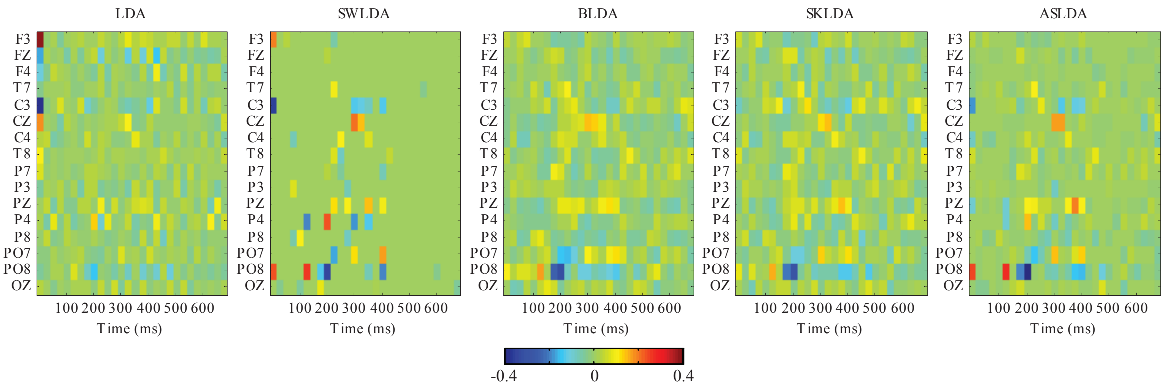

Fig. 7. Discriminant vectors obtained by the LDA, SWLDA, BLDA, SKLDA and ASLDA, respectively, from subject B of the Dataset-1 (using 10 training characters, i.e., 1800 samples), presented in the form of channels × temporal points.

recorded within limited calibration time is insufficient compared to dimensionality of the feature vectors that are constructed by concatenation of multiple temporal points from multiple channels [7,39]. The sparse discriminant vector derived from the proposed ASLDA algorithm implemented automatic feature selection for dimensionality reduction, thereby effectively alleviating the effects of small training sample size on the ERP classification performance. In the ASLDA, the designed aggregation scheme combined multiple sub-classifiers learned with different regularization parameters to effectively overcome the bias-variance tradeoff, which guaranteed good generalization capacity for test data.

According to the ERP characteristics, the learned discriminant vectors were further analyzed to provide an evidence for superiority of the proposed ASLDA over the other compared classification algorithms. As an example, ERP characteristics were derived from 10 training characters data (i.e., 1800 samples) of subject B in the Dataset-1. Fig. 6(a), (b), (c) present the averaged ERP waveforms at electrodes PO8, Cz and Pz, the scalp topographies of EEG amplitude differences between target and nontarget at 200 ms and 330 ms, and the discriminative information, respectively. The discriminative information was evaluated by squared pointwise biserial correlation coefficients with sign (signed r^2-values). The r-value [7] is defined as:

$$r(x) = \frac{\sqrt{N_1 N_2}}{N_1 + N_2} \frac{\mathrm{mean}\{x_i | i \in \mathcal{I}_1\} - \mathrm{mean}\{x_i | i \in \mathcal{I}_2\}}{\mathrm{std}\{x_i | i \in \mathcal{I}_1, \mathcal{I}_2\}}, \tag{14}$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ denote the indices of samples in classes 1 and -1, respectively. The signed r^2-value is equal to $\mathrm{sign}(x) \cdot r(x)^2$. Distribution of the ERP characteristics is relatively sparse and indicates the optimal (i.e., the most discriminative) features around 200 ms (i.e., N200) at PO8 and around 300 ms (i.e., P300) at Cz, and the suboptimal features around 300 ms at Pz. Fig. 7 depicts discriminant vectors (presented in the form of channels × temporal points) obtained by the LDA, SWLDA, BLDA, SKLDA and ASLDA, respectively. The ordinary LDA obtained relatively poorer discriminant vector whose weights were smoothly assigned to the whole feature dimension, and failed to highlight the most discriminative features. Both the BLDA and SKLDA prevented overfitting to the noises and yielded more effective discriminant vectors than that of the ordinary LDA, which however still provided smooth solutions without outstanding feature reduction capability. The stepwise regression with statistical analysis in the SWLDA obtained the most sparse discriminant vector with strong capability for feature reduction, which however failed to capture the suboptimal P300 features at Pz. With the l_1-regularization and aggregation scheme, the relatively sparse discriminant vector learned by the ASLDA captured both the optimal and suboptimal features accurately with good insight into neural mechanism related to ERPs, thereby effectively implementing automatic feature selection for dimensionality reduction and resulting in better overall classification performance than those of the other compared classification algorithms.

Superior classification performance in small training sample size suggested that the proposed ASLDA could be effective to reduce the calibration time for classification of single-trial ERP. For example, from the results shown in Fig. 3, the ASLDA decreased about 6 training characters to achieve the similar classification performance as that of the LDA. That is, about three and a half minutes were saved from the spent calibration time. Kaper et al. [76] reported a single-trial ERP classification accuracy of 64.5 % by using SVM with 2520 training samples that took more than ten minutes for recording. From the results shown in Table 1, the ASLDA yielded a comparable single-trial ERP classification accuracy of 61.5 % with only 320 training samples that takes little longer than one minute for recording. It can be seen that the ASLDA saved much time for the calibration of ERP-based BCI, which could therefore effectively improve the system practicability.

To comprehensively evaluate the performance of proposed algorithm, computational cost of the ASLDA was compared with those of the LDA, SWLDA, BLDA and SKLDA based on the Dataset-3. Under the computation environment of Matlab R2011a on a laptop with 1.20 GHz CPU (3 GB RAM), the computational time was 0.398 s for the LDA, 0.572 s for the SWLDA, 1.823 s for the BLDA, 2.120 s for the SKLDA and 29.15 s for the ASLDA on processing the Dataset-3. Although the ASLDA increased the computational time in contrast to the other algorithms, it is still acceptable compared with the time spent on calibration data recording. Such computational time could be effective reduced by using a more advanced computer system, and hence would not bring significant impact on the BCI system performance. Importantly, the ASLDA achieved better overall classification performance over the other algorithms, which provided relatively higher efficiency for the ERP-based BCI, especially in small sample size scenario.

It is worth noting that the semi-supervised scheme has also recently been proposed for calibration time reduction and adaptive learning of BCI [77–79], and achieved good classification performance, especially with limited training samples. Thus, extension of the proposed ASLDA to a semi-supervised version may improve further the BCI system performance, which will be investigated in our future study.

5. Conclusions

In this study, we introduced an aggregation of sparse linear discriminant analyses (ASLDA) to ERP classification in small sample size scenario for BCI application. Multiple sparse discriminant vectors learned from differently l_1-regularized LSRs were aggregated to result in an ensemble classifier with good strength on automatic feature selection for dimensionality reduction and effective generalization capacity for test data. Extensive experiment analyses were implemented on the dataset II of BCI competition III, the dataset of EPFL BCI group and the dataset recorded from our own BCI experiment. The results show that the proposed ASLDA algorithm yielded better overall performance for single-trial ERP classification in contrast to the ordinary LDA and other competing classification algorithms used in the ERP-based BCIs. Further study will investigate the application of ASLDA in other types of BCIs.

Acknowledgements

The authors sincerely thank the editor and the anonymous reviewers for their insightful comments and suggestions that helped improve the paper. This study was supported in part by the Nation Nature Science Foundation of China under Grant 61305028, Grant 61074113, Grant 61203127, Grant 61103122, Grant 61202155, Fundamental Research Funds for the Central Universities Grant WH1314023, Grant WH1114038, Shanghai Leading Academic Discipline Project B504, and JSPS KAKENHI Grant 24700154.

References

1. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfutscheller and T. M. Vaughan, Brain-computer interfaces for communication and control, *Clin. Neurophysiol.* **113**(6) (2002) 767–791.
2. A. Ortiz-Rosario and H. Adeli, Brain-Computer Interface Technologies: From signal to action, *Rev. Neurosci.* (2013) DOI 10.1515/revneuro-2013-0032.
3. M. A. Lopez-Gordo, F. Pelayo, A. Prieto, and E. Fernandez, An auditory brain-computer interface with accuracy prediction, *Int. J. Neural Syst.*, **22**(3) (2012) 1250009.
4. W. Y. Hsu, Continuous EEG signal analysis for asynchronous BCI application, *Int. J. Neural Syst.* **21**(4) (2011) 335–350.
5. J. Li, J. Liang, Q. Zhao, J. Li, K. Hong and L. Zhang, Design of assistive wheelchair system directly steered

by human thoughts, *Int. J. Neural Syst.* **23**(3) (2013) 1350013.

6. F. Cong, A.-H. Phan, P. Astikainen, Q. Zhao, Q. Wu, J. Hietanen, T. Ristaniemi and A. Cichocki, Multi-domain feature extraction for small event-related ptentials through nonnegative multi-way array decomposition from low dense array EEG, *Int. J. Neural Syst.* **23**(2) (2013) 1350006.
7. B. Blankertz, S. Lemm, M. Treder, S. Haufe and K. R. Müller, Single-trial analysis and classification of ERP components - A tutorial, *NeuroImage* **56**(2) (2011) 814–825.
8. G. Pfurtscheller and C. Neuper, Motor imagery and direct brain-computer communication, *Proc. IEEE* **89**(7) (2001) 1123–1134.
9. C. Park, D. Looney, N. Rehman, A. Ahrabian and D. P. Mandic, Classification of motor imagery BCI using multivariate empirical mode decomposition, *IEEE Trans. Neural Syst. Rehabil. Eng.* **21**(1) (2012) 10–22.
10. X. Gao, D. Xu, M. Cheng and S. Gao, A BCI-based environmental controller for the motion-disabled, *IEEE Trans. Neural Syst. Rehabil. Eng.* **11**(2) (2003) 137–140.
11. N. V. Manyakov, N. Chumerin, and M. M. Van Hulle, Multichannel decoding for phase-coded SSVEP brain-computer interface, *Int. J. Neural Syst.* **22**(5) (2012) 1250022.
12. E. W. Sellers and E. Donchin, A P300-based brain-computer interface: Initial tests by ALS patients, *Clin. Neurophysiol.* **117**(3) (2006) 538–548.
13. B. Blankertz, G. Curio and K. R. Müller, Classifying single trial EEG: Towards brain computer interfacing, *In: Proc. Advances in Neural Information Processing Systems (NIPS 01)* **4** (2002) 157–164.
14. B. Hong, F. Guo, T. Liu, X. Gao and S. Gao, N200-speller using motion-onset visual response, *Clin. Neurophysiol.* **120**(9) (2009) 1658–1666.
15. Y. Zhang, Q. Zhao, J. Jin, X. Wang and A. Cichocki, A novel BCI based on ERP components sensitive to configural processing of human faces, *J. Neural Eng.* **9**(2) (2012) 026018.
16. F. Cong, A.-H. Phan, Q. Zhao, T. Huttunen-Scott, J. Kaartinen, T. Ristaniemi, H. Lyytinen and A. Cichocki, Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from EEG collected by low-density array, *Int. J. Neural Syst.* **22**(6) (2012) 1250025.
17. F. Cong, I. Kalyakin, T. Huttunen-Scott, H. Li, H. Lyytinen, T. Ristaniemi, Single-trial based independent component analysis on mismatch negativity in children, *Int. J. Neural Syst.* **20**(04) (2010) 279–292.
18. Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang and A. Cichocki, Spatial-temporal discriminant analysis for ERP-based brain-computer interface, *IEEE Trans. Neural Syst. Rehabil. Eng.* **21**(2) (2013) 233–243.
19. W. Y. Hsu, Single-trial motor imagery classification using aysmmetry ratio, phase relation, wavelet-based fractal, and their selected combination, *Int. J. Neural Syst.* **23**(2) (2013) 1350007.
20. H. Lee, Y. Kim, A. Cichocki and S. Choi, Nonnegative tensor factorization for continuous EEG classification, *Int. J. Neural Syst.* **17**(4) (2007) 305–317.
21. F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche and B. Arnaldi, A review of classification algorithms for EEG-based brain-computer interfaces, *J. Neural Eng.* **4**(2) (2007) R1–R13.
22. H. Adeli and M. Wu, Regularization neural network for construction cost estimation, *J. Constr. Eng. Manage., ASCE* **124**(1) (1998) 18–24.
23. M. Ahmadlou and H. Adeli, Enhanced probabilistic neural network with local decision circles: A robust classifier, *Integr. Comput-Aided Eng.* **17**(3) (2010) 197–210.
24. H. Adeli, S. Ghosh-Dastidar, N. Dadmehr, A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimers disease, *Neurosci. Lett.* **444**(2) (2008) 190–194.
25. M. Cabrerizo, M. Ayala, M. Goryawala, P. Jayakar and M. Adjouadi, A new parametric feature descriptor for the classification of epileptic and control EEG records in pediatric population, *Int. J. Neural Syst.* **22**(2) (2012) 1250001.
26. U. R. Acharya, S. Vinitha Sree, A. P. C. Alvin and J. S. Suri, Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals, *Int. J. Neural Syst.* **22**(2) (2012) 1250002.
27. Y. Tang and D. M. Durand, A novel electrical stimulation paradigm for the suppression of epileptiform activity in an in vivo model of mesial temporal lobe status epilepticus, *Int. J. Neural Syst.* **22**(3) (2012) 1250006.
28. K. Pauwels and M. M. Van Hulle, Head-centric disparity and epipolar geometry estimation from a population of binocular energy neurons, *Int. J. Neural Syst.* **22**(3) (2012) 1250007.
29. G. Puscasu and B. Codres, Nonlinear system identification and control based on modular neural networks, *Int. J. Neural Syst.* **21**(4) (2011) 319–334.
30. N. M. Tomasevic, A. M. Neskovic, and N. J. Neskovic, Artificial neural network-based approach to EEG signal simulation, *Int. J. Neural Syst.* **22**(3) (2012) 1250008.
31. S. Ghosh-Dastidar and H. Adeli, A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection, *Neural Netw.* **22**(10) (2009) 1419–1431.
32. M. Ahmadlou and H. Adeli, Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD, *Clin. EEG Neurosci.* **41**(1) (2010) 1–10.
33. M. Ahmadlou, H. Adeli and A. Adeli, New diagnostic EEG markers of the Alzheimer's disease using visibility graph, *J. Neural Transm.* **117**(9) (2010) 1099–1109.

34. M. Ahmadlou and H. Adeli, Fuzzy synchronization likelihood with application to attention-deficit/hyperactivity disorder, *Clin. EEG Neurosci.* **42**(1) (2011) 6–13.
35. A. Rakotomamonjy and V. Guigue, BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller, *IEEE Trans. on Biomed. Eng.* **55**(3), (2008) 1147–1154.
36. S. Lemm, B. Blankertz, G. Curio and K. R. Müller, Spatio-spectral filters for improving the classification of single trial EEG, *IEEE Trans. Biomed. Eng.* **52**(9) (2005) 1541–1548.
37. V. Bostanov, BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram, *IEEE Trans. Biomed. Eng.* **51**(6) (2004) 1057–1061.
38. B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio and K. R. Müller, The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naive subjects, *IEEE Trans. Biomed. Eng.* **55**(10) (2008) 2452–2462.
39. U. Hoffmann, J. M. Vesin, T. Ebrahimi and K. Diserens, An efficient P300-based brain-computer interface for disabled subjects, *J. Neurosci. Meth.* **167**(1) (2008) 115–125.
40. J. Jin, B. Z. Allison, E. W. Sellers, C. Brunner, P. Horki, X. Wang and C. Neuper, Optimized stimulus presentation patterns for an event-related potential EEG-based braincomputer interface. *Med. Biol. Eng. Comput.* **49**(2) (2011) 181–191.
41. H. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition, *IEEE Trans. Neural Netw.* **20**(1) (2009) 103–123.
42. P. Schneider, K. Bunte, H. Stiekema, B. Hammer, T. Villmann and M. Biehl, Regularization in matrix relevance learning, *IEEE Trans. Neural Netw.* **21**(5) (2010) 831–839.
43. D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan and J. R. Wolpaw, Toward enhanced P300 speller performance, *J. Neurosci. Meth.* **167**(1) (2008) 15–21.
44. D. J. C. MacKay, Bayesian interpolation, *Neural Comput.* **4**(3) (1992) 415–447.
45. C. Vidaurre, N. Krämer, B. Blankertz and A. Schlögl, Time domain parameters as a feature for EEG-based brain computer interfaces, *Neural Netw.* **22**(9) (2009) 1313–1319.
46. J. Schäfer and K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for function genomics, *Stat. Appl. Genet. Mol. Biol.* **4**(1) (2005) Article 32.
47. G. Rodríguez-Bermudez, P. J. García-Laencina, J. Roca-González and J. Roca-Dorda, Efficient feature selection and linear discrimination of EEG signals, *Neurocomputing* **115**(4) (2013) 161–165.
48. Y. Shin, S. Lee, J. Lee, H.-N. Lee, Sparse representation-based classification scheme for motor imagery-based braincomputer interface systems, *J. Neural Eng.* **9**(5) (2012) 056002.
49. J. Fruitet, D. J. McFarland and J. R. Wolpaw, A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based braincomputer interface, *J. Neural Eng.* **7**(1) (2010) 016003.
50. H. Lu, H. Eng, C. Guan, K. N. Plataniotis and A. N. Venetsanopoulos, Regularized common spatial pattern with aggregation for EEG classification in small-sample setting, *IEEE Trans. Biomed. Eng.* **57**(12) (2010) 2936–2946.
51. R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
52. A. Samant and H. Adeli, Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis, *Comput-Aided Civ. Infrastruct. Eng.* **15**(4) (2000) 241–250.
53. J. Ye, Least squares linear discriminant analysis. *In: Proc. 24th Int'l Conf. Machine learning (ICML 07)* (2007) 1087–1093.
54. S. J. Raudys and A. K. Jain, Small sample size effects in statistical pattern recognition: Recommendations for practitioners, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(3) (1991) 252–264.
55. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, New York: Wiley (1977).
56. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B* **58**(1) (1996) 267–288.
57. H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. B* **67**(2) (2005) 301–320.
58. J. Friedman, T. Hastie and R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Softw.* **33**(1) (2010) 1–22.
59. H. Kang, Y. Nam and S. Choi, Composite common spatial pattern for subject-to-subject transfer, *IEEE Signal Process. Lett.* **16**(8) (2009) 683–686.
60. M. Skurichina and R. P. W. Duin, Bagging, boosting and the random subspace method for linear classifiers, *Pattern Anal. Appl.* **5**(2) (2002) 121–135.
61. J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos and S. Z. Li, Ensemble-based discriminant learning with boosting for face recognition, *IEEE Trans. Neural Netw.* **17**(1) (2009) 166–178.
62. T. Fawcett, An introduction to ROC analysis, *Pattern Recogn. Lett.* **27**(8) (2006) 861–874.
63. D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan and J. R. Wolpaw, A comparison of classification techniques for the P300 speller, *J. Neural Eng.* **3**(4) (2006) 299–305.
64. B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. R. Millan, M. Schroeder and N. Birbaumer, The BCI competition III: Validating alternative approaches to

actual BCI problem, *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**(2) (2006) 153–159.

65. M. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder, *J. Clin. Neurophysiol.* **27**(5) (2010) 328–333.
66. Z. Sankari and H. Adeli, Probabilistic neural networks for diagnosis of Alzheimer's disease using conventional and wavelet coherence, *J. Neurosci. Meth.* **197**(1) (2011) 165–170.
67. Z. Sankari, H. Adeli and A. Adeli, Intrahemispheric, interhemispheric, and distal EEG coherence in Alzheimer's disease, *Clin. Neurophysiol.* **122**(5) (2011) 897–906.
68. M. Ahmadlou and H. Adeli, Functional community analysis of brain: a new approach for EEG-based investigation of the brain pathology, *NeuroImage* **58**(2) (2011) 401–408.
69. M. Ahmadlou, H. Adeli and A. Adeli, Graph theoretical analysis of organization of functional brain networks in ADHD, *Clin. EEG Neurosci.* **43**(1) (2012) 5–13.
70. M. Ahmadlou and H. Adeli, Visibility graph similarity: A new measure of generalized synchronization in coupled dynamic systems, *Physica D* **241**(4) (2012) 326–332.
71. Z. Sankari, H. Adeli and A. Adeli, Wavelet coherence model for diagnosis of Alzheimer disease, *Clin. EEG Neurosci.* **43**(3) (2012) 268–278.
72. M. Ahmadlou, H. Adeli and A. Adeli Improved visibility graph fractality with application for the diagnosis of autism spectrum disorder, *Physica A* **391**(20) (2012) 4720–4726.
73. M. Ahmadlou, H. Adeli and A. Adeli, Fractality analysis of frontal brain in major depressive disorder, *Int. J. Psychophysiol.* **85**(2) (2012) 206–211.
74. M. Ahmadlou, H. Adeli and A. Adeli, Fuzzy synchronization likelihood-wavelet methodology for diagnosis of autism spectrum disorder, *J. Neurosci. Meth.* **211**(2) (2012) 203–209.
75. M. Ahmadlou, H. Adeli and A. Adeli, Spatiotemporal analysis of relative convergence of EEGs reveals differences between brain dynamics of depressive women and men, *Clin. EEG Neurosci.* **44** (2013) 175–181.
76. M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner and H. Ritter, BCI competition 2003 - Data set IIb: Support vector machines for the P300 speller paradigm. *IEEE Trans. Biomed. Eng.* **51**(6) (2004) 1073–1076.
77. Y. Li and C. Guan, An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces, *Neural Comput.* **18**(11) (2006) 2730–2761.
78. Y. Li and C. Guan, Joint feature re-extraction and classification using an iterative semi-supervised support vector machine algorithm, *Mach. Learn.* **71**(1) (2008) 33–53.
79. Y. Li, C. Guan, H. Li and Z. Chin, A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system, *Pattern Recognit. Lett.* **29**(9) (2008) 1285–1294.