Cogn Comput (2014) 6:477–483
DOI 10.1007/s12559-014-9264-1

A Voting Optimized Strategy Based on ELM for Improving Classification of Motor Imagery BCI Data

Lijuan Duan · Hongyan Zhong · Jun Miao · Zhen Yang · Wei Ma · Xuan Zhang

Received: 22 August 2013 / Accepted: 8 April 2014 / Published online: 22 April 2014

Abstract This paper presents an approach to classifying electroencephalogram (EEG) signals for brain–computer interfaces (BCI). To eliminate redundancy in high-dimensional EEG signals and reduce the coupling among different classes of EEG signals, we use principle component analysis and linear discriminant analysis to extract features that represent the raw signals. Next, we introduce the voting-based extreme learning machine to classify the features. Experiments performed on real-world data from the 2003 BCI competition indicate that our classification method outperforms state-of-the-art methods in speed and accuracy.

Keywords Brain–computer interface · Principle component analysis · Linear discriminate analysis · Voting-based extreme learning machine

Introduction

Human brains are natural cognitive systems that have powerful abilities to communicate with each other through external sensing and internal computation. Discovering and simulating the cognitive mechanisms of human brains is always the goal of cognitive computation societies. Although human brains are skull-enclosed, electrical signals at the skull can be detected and analyzed to build artificial cognitive systems that can communicate with the natural cognitive systems. Such artificial cognitive systems are capable of interacting with humans by human–computer interaction (HCI) technologies, such as detecting keywords in human speech [1], predicting human gaze location in dynamic scenes [2], handwritten text recognition [3] and controlling external devices with brain activity in brain–computer interfaces (BCI) [4]. Among these technologies, BCI is noninvasive and allows a brain to directly control a device, bypassing the use of muscular activity. Motor imagery is considered to be mental rehearsal of muscular activity, and various acquisition techniques are available for capturing motor imagery brain activity. Among these techniques, the electroencephalogram (EEG) is most commonly employed for monitoring brain activity in BCI systems. It requires relatively simple and inexpensive equipment and it is more convenient to use than other methods [5].

There are two fundamental stages in the processing of EEG signals: feature extraction and classification [6]. A great variety of features represented on EEG signals have been explored, such as amplitude values [7], band powers (BP) [8], power spectral density (PSD) values [9] and autoregressive (AR) and adaptive autoregressive (AAR) parameters [10]. Classification methods widely used include k-nearest neighbor [7], support vector machines (SVM) [11], neural networks [12], naive Bayes [13], and so on. In this paper, we use LDA-after-PCA to obtain low-dimensional feature representations of original EEG signals. PCA is commonly used in dimension reduction; however, it does not take inter-class differences into account. Thus, features obtained by PCA are not discriminative enough. LDA is capable of obtaining the best projection directions and generates features with maximum

L. Duan · H. Zhong · Z. Yang · W. Ma · X. Zhang
Department of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China

J. Miao (✉)
Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
e-mail: jmiao@ict.ac.cn

between-class scatter and minimum within-class scatter simultaneously [14]. Features generated by the LDA-after-PCA algorithm are theoretically more discriminative than those produced by PCA, which we confirm by extensive experiment in "Evaluation of LDA-after-PCA and V-ELM" section.

The fast and accurate classification of EEG signals is very challenging [15]. Huang et al. [16] proposed a new learning algorithm, called extreme learning machine (ELM), for single hidden layer feed forward neural networks (SLFNs). The algorithm randomly chooses hidden nodes and analytically determines the output weights of SLFNs. It provides good generalization performance and is thousands of times faster than conventional learning algorithms for SLFNs [17]. However, the randomness of the hidden layer parameters of the ELM generally leads to unstable output. Therefore, researchers usually adopt the average of the output of multiple experiments, denoted as average ELM (avg-ELM). More recently, Cao et al. [18] proposed a voting-based ELM (V-ELM) classification method, which incorporates a voting strategy into ELM in classification applications. V-ELM overcomes the instability problem of ELM and obtains better classification accuracy.

In this paper, we present a method to classify EEG signals based on V-ELM. In the first step, effective electrodes of the EEG signal are selected. Next, for fast classification, we use the LDA-after-PCA algorithm [14] to represent the original EEG signals. Finally, we use V-ELM to classify the features. Our EEG classification method (LDA-after-PCA + V-ELM) achieves better performance than state-of-the-art methods. The contributions of this paper mainly include: (1) to the best of our knowledge, we are the first to apply V-ELM to classify EEG signals; (2) a classification method integrating V-ELM and the LDA-after-PCA feature extraction algorithm is proposed; (3) the classification method is shown experimentally to be more accurate than state-of-the-art methods. All the experiments are performed on the BCI competition 2003 data set Ia.

The remainder of this paper is organized as follows. In "Methods" section, we present the proposed classification method, which includes feature extraction and classification. In "Experimental Results" section, we describe the parameter selection for feature extraction and performance evaluation of the classification method, respectively. Finally, the "Discussion and Conclusion" are given.

Methods

The EEG signal data set is divided into two groups for training and testing, respectively. Because of the complexity of EEG signals, it is necessary to extract features to represent them. Hence, we first employ the LDA-after-PCA algorithm to extract a set of features from each sample. During the training process, all the training samples' features are entered into an ELM classifier to learn a classification model using the method in [18]. In all, Z numbers of ELM classifiers are trained based on the same training samples. Each testing sample would have Z labels predicted by the Z ELM classifiers. Its final label is obtained by voting. The classification accuracy is computed by comparing the final voted labels with ground truth.

LDA-After-PCA Feature Extraction

We denote the training data as D, and the testing data as T. The LDA-after-PCA feature extraction algorithm is performed as follows. First, the covariance matrix of all the data is decomposed to obtain eigenvectors and eigenvalues. The basis $M_{PCA} = [\Phi_1, \Phi_2, \ldots, \Phi_m]$ is m eigenvectors corresponding to the top m eigenvalues. A principle component with a large eigenvalue has a large contribution rate to the original data. The sum of the contribution rates of the first m principal components is the accumulative contribution rate (ACR) of the m dimensions of the features. In practice, m is chosen by a given threshold of ACR. Then, we obtain the projected training matrix D_{PCA} and testing matrix T_{PCA}.

$$D_{PCA} = D \cdot M_{PCA} \tag{1}$$

$$T_{PCA} = T \cdot M_{PCA} \tag{2}$$

Next, we apply LDA to D_{PCA} and T_{PCA}, and then obtain LDA remapped matrixes X^1 and X^2. Compared with D_{PCA} and T_{PCA}, X^1 and X^2 have the maximum between-class scatter and minimum within-class scatter.

Classification

After introducing the feature extraction method, we proceed to describe the V-ELM-based classification. We first briefly review ELM, which is the foundation of V-ELM.

Brief Review of ELM

ELM, proposed by Huang et al. [15], has been widely adopted in pattern classification in recent years. It has many advantages, and not only avoids many problems encountered by traditional gradient-based neural network learning algorithms such as local minima and various training parameters (training efficiency, stopping criteria, learning epochs and the hidden layer unit number), but also learns much faster, with higher generalization performance than the established gradient-based learning methods.

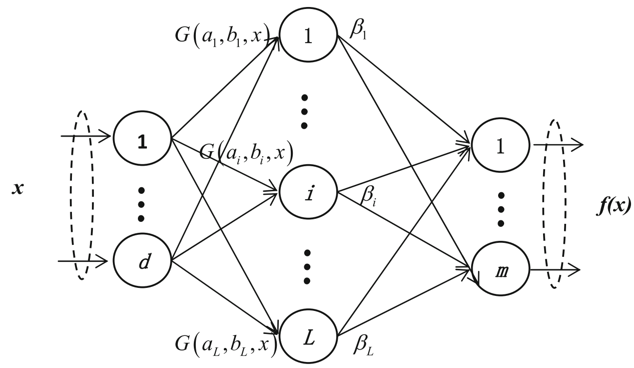

Fig. 1 Single hidden layer feed forward networks

ELM works for generalized SLFNs [19]. The essence of ELM is that the hidden layer of SLFNs need not be tuned. The structure of the SLFNs is shown in Fig. 1.

The output function of SLFNs with L hidden nodes in the output layer can be expressed by Eq. 3.

$$f_L(x) = \sum_{i=1}^{L} \beta_i g_i(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x), \tag{3}$$

where x is the input of ELM, $x \in R^d$; a_i, b_i are the parameters of the hidden layer's nodes; and β_i denotes the linked weights between the i^{th} hidden neurons and output neurons, $\beta_i \in R^m$.

The activation function g, $g_i(x)$ is defined as

$$g_i(x) = G(a_i, b_i, x) = g(x \cdot a_i + b_i), a_i \in R^d, b_i \in R. \tag{4}$$

For RBF nodes with g, $g_i(x)$is defined as

$$g_i(x) = G(a_i, b_i, x) = g(b_i\|x - a_i\|), a_i \in R^d, b_i \in R^+, \tag{5}$$

a_i and b_i denote the center and impact factor of the radial basis function nodes.

For N arbitrary distinct samples, the SLFNs can approximate these N samples with zero error, which implies that there exist β_i, a_i and b_i such that

$$\sum_{i=1}^{L} \beta_i G(a_i, b_i, x_j) = o_j, \quad j = 1, \ldots, N. \tag{6}$$

All the equations above can be written compactly as $G\beta = O$, where

$$G = \begin{bmatrix} G(a_1, b_1, x_1) & \ldots & G(a_L, b_L, x_1) \\ \vdots & \ldots & \vdots \\ G(a_1, b_1, x_N) & \ldots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L}, \tag{7}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \tag{8}$$

$$O = \begin{bmatrix} o_1^T \\ \vdots \\ o_L^T \end{bmatrix}_{L \times m}. \tag{9}$$

O is called the hidden layer output matrix of the SLFN. Training an SLFN is simply equivalent to find a least-squares solution $\overset{\wedge}{\beta}$ of the linear system $G\beta = O$:

$$\left\|G\hat{\beta} - O\right\| = \min_{\beta}\|G\beta - O\|. \tag{10}$$

If $L = N$, SLFNs can approximate these training data with zero error. However, in most cases, the number of hidden nodes is much less than the number of distinct training samples, $L \ll N$. H is a nonsquare matrix and there may not exist a_i, b_i, β_i (i = 1, …, L) such that $G\beta = O$. The smallest norm least-squares solution of the above linear system is as follows:

$$\widehat{\beta} = G^+ O, \tag{11}$$

where G^+ is the Moore–Penrose generalized inverse of matrix G [20, 21]. Thus, ELM can be summarized as follows:

Given training data set$\{(x_i, y_i)\}_{i=1}^{N} \in R^d \times R^m$, activation function $G(a_i, b_i, x)$ and the number of hidden neurons L, the ELM algorithm can be described as follows:

1. Generate the parameters of the hidden nodes randomly;
2. Compute the output matrix G of the hidden layer;
3. Output the optimal weight $\hat{\beta}$ of the network.

Voting Based on ELM (V-ELM)

ELM shows great performance in pattern classification. However, the randomness of the hidden layer parameters leads to unstable output. The most common method of dealing with the instability is to take the average result of multiple running times. We adopted voting based on ELM proposed in [18].

Given training data set $\{x_i^1, y_i^1\}_{i=1}^{N}$, where x_i^1 denotes the *ith* sample of training data, y_i^1 denotes the given target label of the *ith* sample of training data. For binary class problem, $y_i \in \{0, 1\}$. Similarly, for the testing data set $\{x_i^2, y_i^2\}_{i=1}^{F}$, x_i^2 denotes the *i*th sample of testing data, and y_i^2 denotes the given target label of the *i*th sample of testing data.

The details of the strategy are summarized as follows:

1. Train Z ELM classifiers;
2. The *ith* sample of testing data is fed into the *jth* ELM classifier; the predicted label y_{ij} of *ith* sample of testing data set will be outputted.
3. All the testing data are fed into Z numbers of ELMs. Finally, class label set ψ is as follows:

$$\Psi = \begin{bmatrix} y_{11} & \cdots & y_{1Z} \\ \vdots & \cdots & \vdots \\ y_{F1} & \cdots & y_{FZ} \end{bmatrix}_{F \times Z}. \qquad (12)$$

4. For the *ith* testing sample, h^0 is the total number of the ELMs that predict the class of the *ith* sample to be 0; h^1 is the total number of ELMs that predict the class of the *ith* sample to be 1. The sum of h^0 and h^1 is Z.

The final predicted label y_i of the *ith* testing sample can be obtained. The vote function is given by:

$$V(y_i) = \begin{cases} 0 & \text{if} \quad h^0 > h^1 \\ 1 & \text{otherwise } h^0 < h^1 \end{cases}. \qquad (13)$$

Comparing $V(y_i)$ with the ground truth, the final classification accuracy is computed.

One important thing to note here is that if Z is an even number, the votes of class 0 for the *ith* sample are equal to the votes of class 1 for the *ith* sample, which will increase the difficulty of making a decision about which class the *ith* sample belongs to. To avoid this case, we train an additional ELM classifier when $h^0 = h^1$ and regard its predicted label as the predicted label of the *ith* sample.

Experimental Results

We use the BCI competition 2003 data set Ia to validate our method. In this section, we first introduce the data set. Then, we discuss parameter selection in feature extraction and V-ELM. Finally, we evaluate the proposed method in accuracy and computation time, by comparing it with combinations of different features and different classification methods on the same data set.

Data Set Description

BCI competition 2003 data set Ia was taken from a healthy participant [22]. The participant was asked to move a cursor up (class 0) and down (class 1) on a computer screen while his slow cortical potentials (SCPs) were taken. During the recording, the participant received visual feedback of his SCPs. Cortical positivity led to a downward movement of the cursor on the screen, while cortical negativity led to an upward movement. Brain activity was recorded from six EEG channels sampled at 256 Hz. Six EEG electrodes were located according to the International 10–20 system as shown in Fig. 2. All the trials were from the training set (268 trials, 135 for class 0, 133 for class 1) and testing set (293 trials, 147 for class 0, 146 for class 1). Each trial lasted 6 s. The visual feedback was presented from second 2 to second 5.5. Only these 3.5-s intervals of each trial are included in the data set.

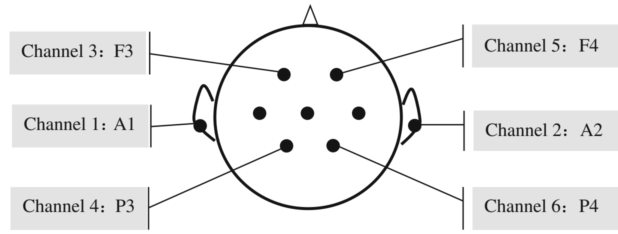

Fig. 2 Distribution of six EEG electrodes

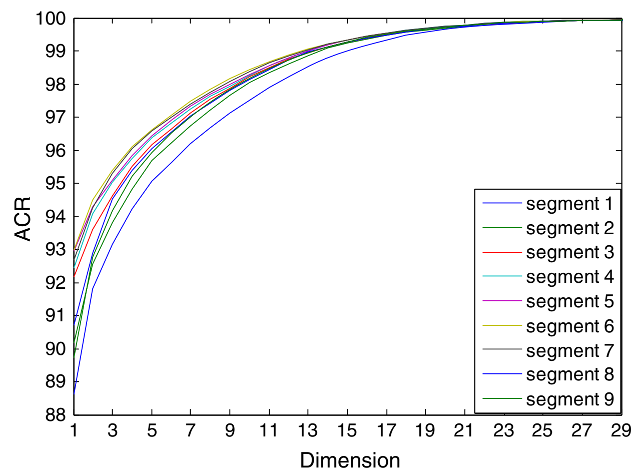

Fig. 3 ACR based on different PCA dimensional reduction

Based on our previous work [23], only signals from channel 1 and channel 2 (A1 and A2) were retained to yield discriminative feature sets. We partitioned the continuous recording samples into subepochs of 500 ms with 125-ms overlap. As a result, the raw data were split into nine segments. Each segment has 128 dimensions.

Parameter Selection of Feature Extraction

LDA-after-PCA is carried out to extract features for imaginary movement sources. First, PCA is applied to each of the nine segments to produce a set of features of s dimensions, separately. We choose $s = 19$, which has an ACR of 99 %, as we can see from Fig. 3. Second, LDA is performed on the features. For binary-class problems, each feature segment is converted to one dimension after LDA. Finally, we combine each segment's LDA features of channel A1 and A2 to form a new feature of 18 dimensions as the input of V-ELM.

Parameter Selection of V-ELM

The optimal number of ELM in our V-ELM classifier is determined as follows. We carry out our V-ELM classification 100 times, with increasing numbers of ELMs,

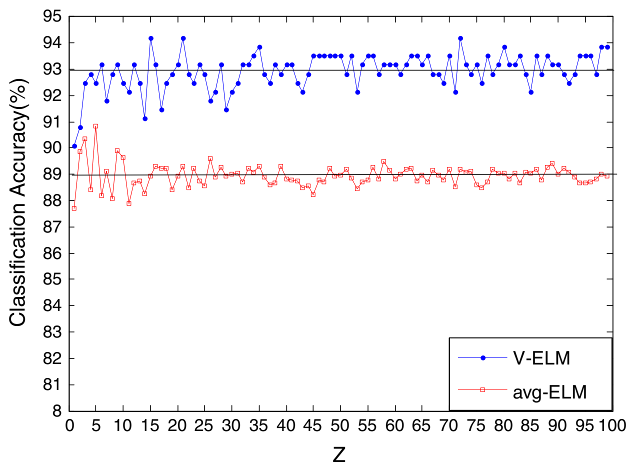

Fig. 4 Classification accuracy of V-ELM and avg-ELM based on the same parameters

denoted as *Z*, ranging from 1 to 100. The classification accuracies of the experiments are given in Fig. 4. From the figure, we observe that the accuracies tend to be stable when Z is larger than 50. To balance the computational complexity and the stability, we set the optimal number of ELMs in our classifier to be 50. Both V-ELM and avg-ELM have 20 hidden nodes. Their activation functions are sigmoid.

Evaluation of LDA-After-PCA and V-ELM

To demonstrate its effectiveness, we compare LDA-after-PCA with PCA and LDA sequentially, together with different types of classifiers. Regularized avg-ELM is the average accuracy of 50 experiments based on regularized ELM [24]. Similar to V-ELM, regularized V-ELM is voting-based regularized ELM.

We compare our classification method with avg-ELM (taking features from LDA-after-PCA) from a macroscopic view. As shown in Fig. 4, avg-ELM becomes stable when the number of ELMs is larger than around 20. The figure demonstrates that the V-ELM classifier has a higher accuracy than avg-ELM, no matter how many ELMs are involved. In the following experiments, we choose the same number of ELMs in avg-ELM with V-ELM.

Table 1 presents the classification accuracy of the combinations. The classification accuracy of PCA + V-ELM is 89.42 %, which is 6.53 % higher than PCA + avg-ELM. The classification accuracy of LDA + V-ELM is 84.30, 4.05 % higher than LDA + avg-ELM. The classification accuracy of LDA-after-PCA + V-ELM is 93.52 %, while the LDA-after-PCA + avg-ELM is only 88.89 %. Therefore, V-ELM outperforms avg-ELM in accuracy. We also try regularized ELM and kernel ELM.

Table 1 Evaluation of LDA-after-PCA and V-ELM

Feature extraction	avg-ELM	V-ELM	Regularized avg-ELM	Regularized V-ELM	Kernel ELM
PCA	82.89	89.42	84.5	89.76	63.48
LDA	80.25	84.30	81.34	84.30	75.09
LDA-after-PCA	88.89	**93.52**	88.83	93.17	92.15

Bold value indicate the best results of our method

Their classification results are also presented in Table 1. From the table, we can see that for all the classifiers, their classification accuracies are highest when choosing LDA-after-PCA. As seen from Table 1, no matter what feature is chosen, the accuracy of voting is always better than the average. V-ELM is better than regularized V-ELM. Therefore, the LDA-after-PCA combining V-ELM method is the optimal schema.

Comparison with Other Classification Methods

We also compare our method with those proposed in [25], which use the same data set. Table 2 lists all the methods, their number of electrodes, feature extraction methods, classifiers and classification accuracies. The classification accuracies are given in the last column. The results indicate that our method has an accuracy of 93.52 %, which is 1.37 % higher than Kayikcioglu and Aydemir's method [25], the best of the others.

We also evaluate our method in terms of computation time. The computation time includes both training and testing time. The classification accuracies and time costs are given in Table 3. As seen from Table 3, the highest accuracy is achieved by V-ELM, which is 0.75 % higher than SVM. V-ELM outperforms SVM both in accuracy and in speed. V-ELM is comparable with avg-ELM in speed, but is much faster than SVM. All the experiments are conducted on a PC with an Intel Core 2 Duo E6550 processor at 2.33 GHz and coded with MATLAB 2012b.

Discussion and Conclusion

In this paper, we presented a classification method, which uses LDA-after-PCA to extract features and V-ELM as the classifier. The main conclusions are summarized as follows:

1. The method is suitable for binary-class signals of motor imagery. The performance of our approach was demonstrated with the BCI completion 2003 data set Ia.
2. Compared with average accuracies, accuracies adopting a voting strategy are more competitive. Accuracies

Table 2 Classification results comparing the proposed method with other methods

Methods	Electrode numbers	Feature extraction	Classifier	Accuracy (%)
Mensh et al. [26]	4	Gamma-band power combined with SCP	Linear	88.70
Wang et al. [27]	2	SCP and beta-band-specific energy	Neural network	91.47
Wu et al. [28]	6	Wavelet package	Neural network	90.80
Sun et al. [29]	6	Combing SCP with the spectral centroid	Bayes	90.44
Kayikcioglu et al. [25]	1	Coefficients of the second order polynomial	k-NN	92.15
Proposed method	2	LDA-after-PCA	V-ELM	**93.52**

Bold value indicate the best results of our method

Table 3 Performance comparison of SVM and V-ELM

Methods	Accuracy (%)	Time cost (s)
SVM	92.15	2.83
V-ELM	**93.52**	**0.33**
Avg-ELM	88.92	0.31

Bold values indicate the best results of our method

using V-ELM are better than using regularized V-ELM. LDA-after-PCA is the most efficient feature extraction method. We combined LDA-after-PCA and V-ELM as the EEG signal processing method and applied it to the BCI competition 2003 data set Ia. Experimental results showed that: (1) it can generate better classification performance than state-of-the-art methods; (2) with the same time cost of avg-ELM, V-ELM is better than SVM in both accuracy and time cost.

3. We also used the method on data set Ib, on which Bostanov obtained the best accuracy result of 54.4 % in the BCI 2003 competition [30]. Wu et al. [29] achieved an accuracy of 59.1 %, and Kayikcioglu et al. [25] obtained 58.9 %. In our study, we achieved 55.56 % with our feature extraction and classification method. As mentioned in Ref. [25], the reported results, including ours, are poor, which may indicate that this data set may not contain task-related information.
4. Although only the binary-class classification strategy is discussed, V-ELM can also be designed to solve the multi-classification problem. We believe that this method has great potential for the design of real-time systems.

Acknowledgments This research is partially sponsored by the Natural Science Foundation of China (Nos. 61175115, 61272320, 61001178, 61003105 and 61370113), the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (CIT&TCD201304035, CIT&TCD201404052), the Jing-Hua Talents Project of Beijing University of Technology (2014-JH-L06), the Ri-Xin Talents Project of Beijing University of Technology (2014-RX-L06) and the International Communication Ability Development Plan for Young Teachers of Beijing University of Technology (No. 2014-16).

References

1. Wöllmer M, Eyben F, Graves A, Schuller B, Rigoll G. Bidirectional LSTM networks for context-sensitive keyword detection in a cognitive virtual agent framework. Cognit Comput. 2010;2(3): 180–90.
2. Mital P, Smith T, Hill R, Henderson J. Clustering of gaze during dynamic scene viewing is predicted by motion. Cognit Comput. 2011;3(1):5–24.
3. Wang Q-F, Cambria E, Liu C-L, Hussain A. Common sense knowledge for handwritten Chinese recognition. Cognit Comput. 2013;5(2):234–42.
4. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM. Brain–computer interface technology: a review of the first international meeting. IEEE Trans Rehabil Eng. 2000; 8(2):164–73.
5. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clin Neurophysiol. 2002;113(6):767–91.
6. Vallabhaneni A, Wang T, He B. Brain–computer interface, neural engineering. US: Springer; 2005. p. 85–121.
7. Kaper M, Meinicke P, Grossekathoefer U, Lingner T, Ritter H. BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. IEEE Trans Biomed Eng. 2004;51(6): 1073–6.
8. Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. EEG-based discrimination between imagination of right and left hand movement. Electroencephalogr Clin Neurophysiol. 1997;103(6): 642–51.
9. Chiappa, S, Bengio S. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. In: ESANN. 2004. p. 193–204.
10. Pfurtscheller G, Neuper C, Schlogl A, Lugger K. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. IEEE Trans Rehabil Eng. 1998;6(3):316–25.
11. Camps-Valls G, Bruzzone L. Kernel-based methods for hyper spectral image classification. IEEE Trans Geosci Remote Sens. 2005;43(6):1351–62.
12. Subasi A, Erçelebi E. Classification of EEG signals using neural network and logistic regression. Comput Methods Programs Biomed. 2005;78(2):87–99.
13. LeVan P, Urrestarazu E, Gotman J. A system for automatic artifact removal in ictal scalp EEG based on independent

component analysis and Bayesian classification. Clin Neurophysiol. 2006;117(4):912–27.
14. Panahi N, Shayesteh MG, Mihandoost S, Varghahan BZ. Recognition of different datasets using PCA, LDA, and various classifiers. In: Proceedings of the 5th international conference on application of information and communication technologies. 2011. 1–5.
15. Song Y-D, Liò P. Epileptic EEG detection via a novel pattern recognition framework. In: Proceedings of the 5th international conference on bioinformatics and biomedical engineering. 2011. 1–6.
16. Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: a new learning scheme of feed forward neural networks. In: Proceedings of IEEE international joint conference on neural networks. 2004. Vol. 2, p. 985–990.
17. Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: theory and applications. Neurocomputing. 2006;70(1):489–501.
18. Cao J, Lin Z, Huang G-B, Liu N. Voting based extreme learning machine. Inf Sci. 2012;185(1):66–77.
19. Huang G-B, Wang D-H, Lan Y. Extreme learning machines: a survey. Int J Mach Learn Cybern. 2011;2(2):107–22.
20. Serre D. Matrices: theory and applications. New York: Springer; 2002.
21. Rao CR, Mitra SK. Generalized inverse of matrices and its applications. New York: Wiley; 1971.
22. Blankertz B, Muller KR, Curio G, Vaughan TM, Schalk G, Wolpaw JR, Schlogl A, Neuper C, Pfurtscheller G, Hinterberger T, Schroder M, Birbaumer N. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Trans Biomed Eng. 2004;51(6):1044–51.
23. Duan L, Zhang Q, Yang Z, Miao J. Research on heuristic feature extraction and classification of EEG signal based on BCI Data Set. Res J Appl Sci Eng Technol. 2013;5(3):1008–14.
24. Deng W, Zheng Q, Chen L. Regularized extreme learning machine: computational Intelligence and data mining, CIDM'09. IEEE Symp. 2009;2009:389–95.
25. Kayikcioglu T, Aydemir O. A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data. Pattern Recognit Lett. 2010;31(11):1207–15.
26. Mensh BD, Werfel J, Seung HS. BCI competition 2003-data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. IEEE Trans Biomed Eng. 2004;51(6):1052–6.
27. Wang B-J, Jun L, Bai J, Peng L, Li Y, Li G. EEG recognition based on multiple types of information by using wavelet packet transform and neural networks, engineering in medicine and biology society, 2005. IEEE-EMBS 2005. 27th annual international conference of the IEEE. 2006: 5377–5380.
28. Sun S, Zhang C. Assessing features for electroencephalographic signal categorization, acoustics, speech, and signal processing, 2005. Proceedings (ICASSP'05) IEEE international conference on IEEE, 5: v/417–v/420 Vol. 5, 2005.
29. Wu T, Yan G-Z, Yang B-H, Sun H. EEG feature extraction based on wavelet packet decomposition for brain computer interface. Measurement. 2008;41(6):618–25.
30. Bostanov V. 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. 2004;51(6):1057–61.