Original Article

Discriminative Common Spatial Pattern Sub-bands Weighting Based on Distinction Sensitive Learning Vector Quantization Method in Motor Imagery Based Brain-computer Interface

Fatemeh Jamaloo, Mohammad Mikaeili

Department of Engineering, Shahed University, Tehran, Iran

ABSTRACT

Common spatial pattern (CSP) is a method commonly used to enhance the effects of event-related desynchronization and event-related synchronization present in multichannel electroencephalogram-based brain-computer interface (BCI) systems. In the present study, a novel CSP sub-band feature selection has been proposed based on the discriminative information of the features. Besides, a distinction sensitive learning vector quantization based weighting of the selected features has been considered. Finally, after the classification of the weighted features using a support vector machine classifier, the performance of the suggested method has been compared with the existing methods based on frequency band selection, on the same BCI competitions datasets. The results show that the proposed method yields superior results on "ay" subject dataset compared against existing approaches such as sub-band CSP, filter bank CSP (FBCSP), discriminative FBCSP, and sliding window discriminative CSP.

Key words: *Brain-computer Interfaces, computer-assisted, electroencephalography, learning, signal processing, support vector machines*

INTRODUCTION

Brain-computer interface (BCI) is a communication channel, which translates the brain activities into the control commands to make it possible for the people with severe neuromuscular disorders to communicate with a computer. Previous studies have shown that imagining the movement of different parts of the body can result in attenuation and enhancement of electroencephalogram (EEG) in corresponding cortex locations and other cortex locations, respectively.[1] These attenuations and enhancements are called event-related desynchronization (ERD) and event-related synchronization (ERS), respectively.

Common spatial pattern (CSP) algorithm has proved to be a highly successful spatial filter for detecting ERD/ERS effects on the motor imagery-based BCI systems. It seeks the feature space for directions that maximize variance for one class and minimize it for the other classes simultaneously.[2] In general, applying CSP to unfiltered or poorly filtered EEG signals will result in poor recognition accuracy.[3] To overcome this limitation of CSP and to select most discriminative frequency bands for the CSP, several approaches have already been proposed – e.g., common spatio-spectral pattern (CSSP),[4] common sparse spectral spatial pattern (CSSSP),[5] sub-band CSP (SBCSP),[3] filter bank CSP (FBCSP),[6,7] discriminative FBCSP (DFBCSP)[8] and sliding window discriminative CSP (SWDCSP),[9] spectrally-weighted CSP,[10] and iterative spatio-spectral pattern learning.[11]

CSSP algorithm utilizes the method of delay embedding in order to extend the CSP algorithm to the state space.[4] CSSSP algorithm allows simultaneous optimization of a spatial and a spectral filter enhancing discriminability rates of multichannel EEG single trials.[5] However, the CSSSP needs extensive parameter tuning and the complexity of the optimization problem leads to computational inefficiency.[1] In SBCSP, the EEG signals are decomposed into sub-bands using a filter bank, and then a score is computed for the features of all sub-bands used in the classification.[3] FBCSP employs a feature selection algorithm to select discriminative CSP features based on the mutual information between the CSP features and the class labels.[6] The DFBCSP selects subject specific discriminative frequency bands using Fisher ratio of filtered EEG signal from channels C3 or C4 in the International 10–20 EEG recording system.[8] To solve the problem of frequency band selection, SWDCSP extracts the CSP features in overlapping frequency bands, and an unsupervised learning algorithm called affinity propagation is used to select a discriminative feature set.[9]

Address for correspondence:
Dr. Fatemeh Jamaloo, Street No. 9, Emarat Street, Shahre Qhods, Tehran, Iran. E-mail: Fatemeh_jamaloo@yahoo.com

In this paper, a novel CSP sub-band feature selection has been provided based on the discriminative information of the features. For this purpose, two factors are considered. The first measure is the mean difference of variances of the first and the last rows of spatially filtered EEG in each sub-band. The second measure is the number of times in all training examples that a sub-band has the maximum difference of variances of the first and the last rows of spatially filtered EEG between all sub-bands.

After the selection of the optimal sub-bands, the distinction sensitive learning vector quantization (DSLVQ) method is used for weighting the selected sub-bands. DSLVQ is a modification of LVQ method, which seeks the most distinctive features on the go. For this purpose, a trainable weight has been assigned to each feature as a measure of importance[12] and these weights have been updated throughout the LVQ iterations. The performance of this method has been evaluated using a support vector machine (SVM) classifier with radial basis function (RBF) kernel. SVM is a supervised machine learning algorithm, which separates the classes with a hyperplane, which is constructed by the observed examples in a higher dimension space.

The remainder of the paper is organized as follows: In Section 2, a short description of CSP, DSLVQ, and SVM algorithms is provided. The dataset and the performed experiments are explained in Section 3. Section 4 discusses the results from the BCI competition III, IVa datasets, and then the performance of the proposed approach is evaluated. Section 5 concludes the paper.

MATERIALS AND METHODS

Common Spatial Pattern

CSP algorithm is highly successful in calculating spatial filters for detection of ERD effects.[13] It is based on the simultaneous diagonalization of two covariance matrices.[9]

Let $R_a(i)$ and $R_b(i)$ denote the corresponding normalized spatial covariance matrices of the trial i from classes a and b, respectively.

$$R_a(i) = X(i)X^T(i) \qquad i = 1:n_1$$
$$R_b(i) = X(i)X^T(i) \qquad i = 1:n_2 \tag{1}$$

where n_1 and n_2 are the number of trials in each class and $X(i)$ is the EEG data of trial i. $X(i)$ is an $N \times T$ matrix, where N is the number of channels and T is the number of samples in time. The averaged normalized covariance matrices over trials are:

$$R_a = \frac{1}{n_1} \sum_{i=1}^{n_1} R_a(i)$$
$$R_b = \frac{1}{n_2} \sum_{i=1}^{n_2} R_b(i) \tag{2}$$

where n_1 and n_2 are the number of trials in each class.

The composite covariance matrix R is defined as follows:

$$R = R_a + R_b \tag{3}$$

Since R is a symmetric matrix, it can be factored as:

$$R = U_0 \Lambda_c U_0^T \tag{4}$$

where U_0 consists of the eigenvectors of R and Λ_c is a diagonal matrix with eigenvalues of R on its main diagonal and U_0^T is the transpose of U_0.

The whitening transformation of the composite covariance matrix R is obtained as follows:

$$P = \Lambda_c^{-\frac{1}{2}} U_0^T \tag{5}$$

Transforming the individual covariance matrices R_a and R_b by P gives:

$$S_a = PR_a P^T$$
$$S_b = PR_b P^T \tag{6}$$

It can easily be shown that S_a and S_b share the same eigenvectors. Thus,

$$S_a = U\Psi_a U^T$$
$$S_b = U\Psi_b U^T \tag{7}$$

From $S_a + S_b = I$ it follows that $\Psi_a + \Psi_b = I$. This means that the eigenvectors with the largest eigenvalue for S_a has the smallest eigenvalue for S_b and vice versa. Throughout this paper, it has been assumed that eigenvalues are sorted in descending order. Note that the projection given by the Pth row of matrix U has a relative variance equal to the Pth element of Ψ_a for trials of class "a" and relative variance equal to the Pth element of Ψ_b for trials of class "b." Since S_a and S_b share the same eigenvectors, by the projection matrix $W_{CSP} = U^T P$, we get the mapping of each EEG trial $X(i)$ as follows:

$$Z(i) = W_{CSP} X(i) \tag{8}$$

The columns of W_{CSP}^{-1} are the CSPs and can be seen as time invariant EEG source distribution vectors. The components most suitable for discrimination are the first and the last few rows of Z.

Finally, the variances of the CSP-based spatially filtered EEG are calculated as the features for the classification task. For the first r and the last r rows of Z, the formula is:

$$f^p(i) = \text{var}(Z^p(i))$$
$$p = 1,...,r,n-r+1,n-r+2,...,n \tag{9}$$

where $Z^p(i)$ signifies the Pth row of the Z(i) and n is the number of rows of Z. In this article, we choose the first and

the last rows of Z as suggested.[3,6,8] It means that we have two features for each trial.

Sub-Band Selection

For the selection of the optimal sub-bands, the frequency band is first divided into equal sub-bands using a filter bank, and then the CSP features for each sub-band is calculated. In order to select the optimal sub-bands from the calculated CSP features in all sub-bands, two measures are considered. The first measure is the average difference between the variances (features) on all trials of each class. Obviously, if point clouds in the scatter plot of features are tinier and longer, the difference of the largest and the smallest variances is more and, in this case, two classes have the most discrepancy. Therefore, the sub-band with a maximum difference in both classes is the best one.

For calculation of the first measure, for each class trial, the mean of differences between the variances of the first and the last rows of spatially filtered EEG is calculated for each sub-band. We define:

$$f_{ak}^p(i) = \mathrm{var}(Z_{ak}^p(i))$$
$$f_{bk}^p(i) = \mathrm{var}(Z_{bk}^p(i)) \qquad P = 1, N \tag{10}$$

where Z_{ak} and Z_{bk} are the CSP-based spatially filtered EEG in the kth sub-band of the first and second class examples, respectively, and N is the last row of Z_{ak} and Z_{bk}. The mean of the variance differences is defined as:

$$d_{ak} = mean(\sum_{i=1}^{N} f_{ak}^1(i) - f_{ak}^n(i))$$

$$d_{bk} = mean(\sum_{i=1}^{N} f_{bk}^1(i) - f_{bk}^n(i)) \tag{11}$$

Finally, the difference between these two values for each sub-band is considered as the first measure. In other words, the sub-band, which includes the greatest mean of variance difference, is the best band.

Concerning the second measure, the difference between the variances of the first and the last rows for each training set is calculated for all sub-bands and sorted in descending order. The sub-band with the maximum difference in most of training examples is considered to be the best one.

Finally, the optimal sub-bands are selected based on these two measures, that is, the sub-band, which maximizes both measures is selected as the best one.

Distinction Sensitive Learning Vector Quantization

DSLVQ method finds the most distinctive features playing a critical role in the success of classification. For this purpose, a weight is assigned to each feature as a scalar indication of the measure of importance and the optimal weight is estimated for each dimension through an iterative learning process. In this context, the amplitude of the weights indicates how informative the features are.[12] Application of the most informative features individually or as a combination produces a significant effect on the classification accuracy.

The learning algorithm for weights is:

$$W(t+1) = norm(W(t) + \beta(t)(norm(W_{new}(t) - W(t))) \tag{12}$$

where $W(t)$ is the present weight vector and $W_{new}(t)$ is the new weight vector during iteration t. Through the learning rate $\beta(t)$, the present weight vector is shifted toward the new vector to some extent. Being smaller than 1, $\beta(t)$ provides a rational decision between the current and the new weight vector during iteration t.

It is important to note that in all LVQ methods, the subset of the feature space points assigned to each class and supposed to represent the distribution of the whole data set in that class is called the codebook vectors, and so the new entry points can be classified according to the class of the closest codebook vectors via the nearest neighbor rule and the Euclidean distance metric.

In the calculation of the new weight vector, we use:

$$W_{new} = \frac{d_o - d_c}{\sum_{i=1}^{n} |d_o(i) - d_{c(i)}|} \tag{13}$$

where n is the number of features for each example and d_c denotes the distance between the features of training example $X(t)$ and the closest codebook from the correct class. Moreover, d_o denotes the distance between $X(t)$ and the closest vector of the other class. If the closest codebook vector to the corresponding feature of training example is from the correct class, then the feature is appropriate for the classification and $W(t+1)$ increases, otherwise it decreases. We have used the DSLVQ method for weighting the CSP features in the selected sub-bands.

Support Vector Machine

A SVM is a learning method developed based on the Vapnik–Chervonenkis dimension theory and structural risk minimization principle in the statistical learning theory. It has good advantages such as strong adaptability, global optimization, high training efficiency, good generalization performance, and small sample learning problem.[13]

Training a SVM is equivalent to solve a quadratic programming problem. Its goal is to find an optimal

separating hyperplane for a given feature set. A classifier implementing the optimal separating hyperplane in the feature space is given by:

$$f(x) = sign\left[\sum_{i=}^{n} y_i a_i K(x, x_i) + e \right]$$ (14)

$$0 \leq a_i \leq c$$

where x_i's are support vectors, y_i's are the labels of x_i's (1 or -1 for a two-class problem), a_i's are Lagrange multipliers, e is classification threshold, $K(.,.)$ is a kernel function, and c is a marginal factor. In this paper, the classification is carried out using SVM with the RBF kernel function:

$$K(x, x_i) = \exp(-r\|x - x_i\|^2)$$ (15)

where r is the width of the RBF.

CALCULATION

Data Description

Our proposed method has been evaluated using the BCI competition III, IVa dataset. This dataset comprises 280 trials of EEG recordings from 118 electrodes for five subjects – namely "aa," "al," "av," "aw," and "ay" – which performs right hand and right foot imagination. Each class has 140 trials. The classification error rate was calculated using 10×10-fold cross-validation on the dataset and compared with the SBCSP, FBCSP, DFBCSP, and SWDCSP results.[3,9]

Data Processing

Data were filtered from 0.5 s to 2.5 s after the visual cue.[3,6,8,9] The EEG data were band-pass filtered using six fifth order Butterworth filters each with a bandwidth of 4 Hz, dividing the frequency range of 8–32 Hz into six equal sub-bands. The filtered EEG data were spatially filtered using the projection matrix of the CSP (W_{CSP}). As mentioned in Section 2, the variances of the spatially filtered signals in the first and last rows are utilized as the features for selection of the optimal sub-bands.

The best two sub-bands were then selected using the proposed frequency band selection method. For further illustration, a comparison of the sub-band selection measures for subject "aa" is provided in Table 1. As it can be observed, the second and the fifth sub-bands are the sub-bands maximizing both measures for band selection.

After the selection of these two optimal sub-bands, they are weighted using DSLVQ method. The number of codebook vectors for most of the subjects were set to six and their initial value in each sub-band was chosen equal to the best sub-band in all the training examples based on the maximum difference of variances between the first and the last rows of the CSP filtered

EEG. Learning rate of the codebook vectors was selected by trial and error in the range of 0.001–0.2 for each subject using grid search and the learning rate of the weights was chosen as 0.1 times the learning rate of the reference vectors.

Finally, the weighted CSP filtered EEG data in the selected sub-bands were classified using an SVM classifier with RBF kernel. The parameters of the classifier were selected in the range of 2^{-15}–2^{15} by logarithmic grid search using cross-validation on the training data set.[14]

RESULTS AND DISCUSSION

For each subject, 10×10-fold cross-validation has been performed on the selected and weighted variances of the spatially filtered EEG signal, and a comparison of the resulting error rate for all subjects is summarized in Table 2. The results show that the proposed method has been shown to yield superior results on "ay" subject dataset compared against existing approaches such as SBCSP, FBCSP, DFBCSP, and SWDCSP.

For further clarification of the properties associated with our proposed method, Figures 1 and 2 show the variances of the training EEG for subject "aa" after projection onto the most

Table 1: Comparison of sub-band selection measures for the sub-bands of subject "aa"

Sub-band	8-12 Hz	12-16 Hz	16-20 Hz	20-24 Hz	24-28 Hz	28-32 Hz
First measure	1.27	1.37	1.02	1.06	1.2	1.03
Second measure	34	54	49	31	73	39

Table 2: Performance comparison of two methods applied on dataset IVa, BCI competition III

Sub-band selection method	Subject "aa"	Subject "al"	Subject "av"	Subject "aw"	Subject "ay"
SBCSP	10.7	1.4	29.6	4.3	4.3
FBCSP	6.9	0.9	31	4.9	6.2
DFBCSP	9.8	1.3	22.2	2.1	5.8
SWDCSP	7	0.1	7.6	2	3.2
Proposed method	7.2	1.4	39	6.1	1.1

SBCSP – Sub-band Common Spatial Pattern; FBCSP – Filter Bank Common Spatial Pattern; DFBCSP – Discriminative Filter Bank Common Spatial Pattern; SWDCSP – Sliding Window Discriminative Common Spatial Pattern; BCI – Brain-computer interface

Table 3: Classification accuracy comparison using DSLVQ for weighting and without using DSLVQ

Subject	Selected sub-bands	With DSLVQ accuracy	Without DSLVQ accuracy
aa	2.5	85.7	85.7
al	1.2	96.4	96.4
av	2.3	56.4	52.1
aw	1.2	81.4	64.3
ay	1.3	83.6	82.1

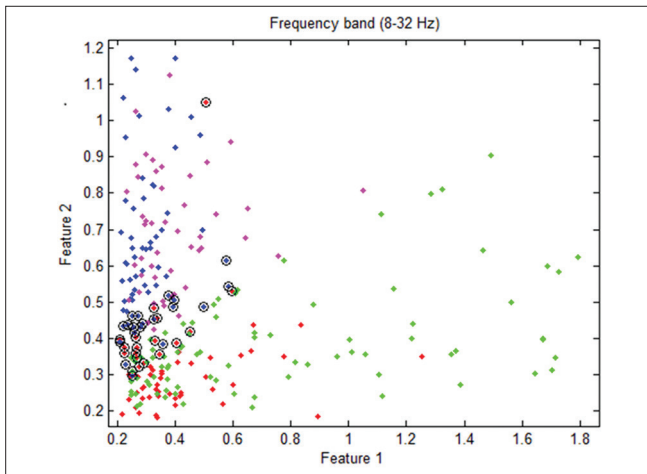DSLVQ – Distinction sensitive learning vector quantization

Figure 1: Variances of the training electroencephalogram signals for subject "aa" after projection onto the most discriminative pairs of directions obtained by basic common spatial pattern. In this figure, blue and red points are training examples and black circles are support vectors of the classifier and green and magnet points are test examples for each class
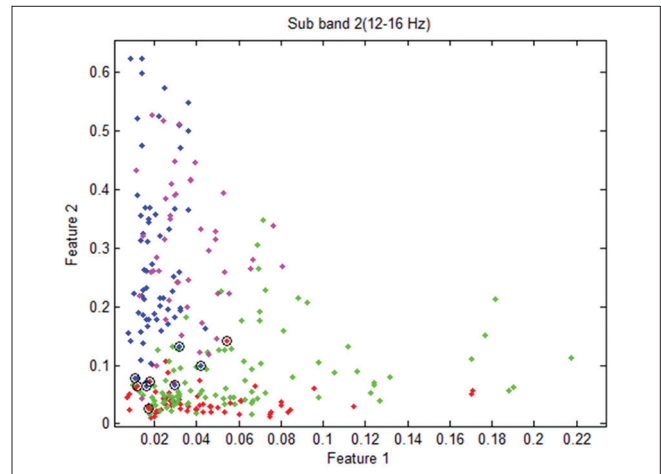

Figure 2: Variances of the training electroencephalogram signals for subject "aa" after projection onto the most discriminative pairs of directions for the second sub-band (12–16 Hz) obtained by sub-band common spatial pattern. In this figure, blue and red points are training examples and black circles are support vectors of the classifier and green and magnet points are test examples for each class
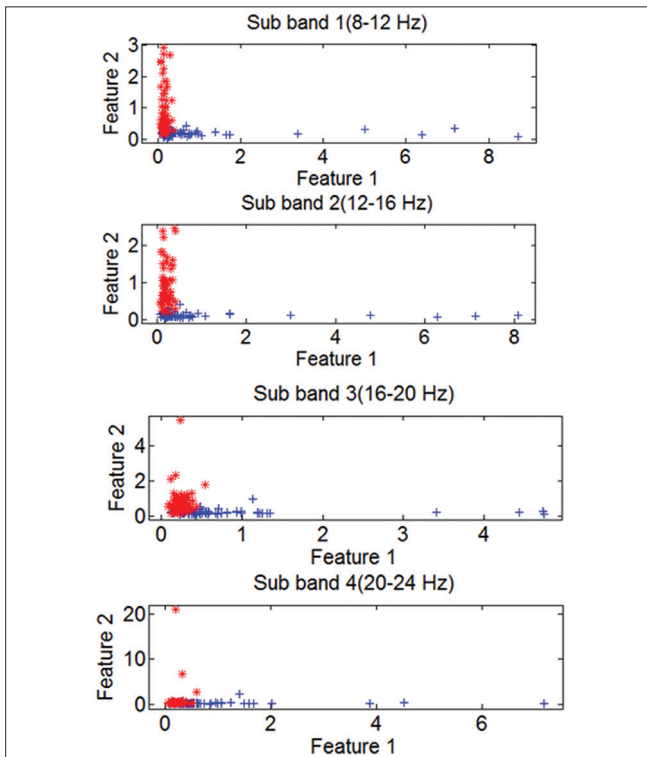

Figure 3: Variances of the training electroencephalogram signals after projection onto the most discriminative pairs of directions for all sub-band obtained by sub-band common spatial pattern for subject "aa"

important discriminative pairs of directions obtained from the basic CSP and from our method, respectively. It means that for each point in these figures, the horizontal coordinate is the variance of the first row of the spatially filtered trial, and the vertical one is the variance of the last row.

The points scattered in Figure 2 is tinier and longer than that in Figure 1, meaning that the difference of the largest

and the smallest variances in the selected sub-bands is greater than that of the other sub-bands and so the features in this sub-band is more discriminative.

Figure 3 shows the variance of the training EEG after projection onto the most important discriminative pairs of directions for all sub-bands for subject "aa." As we can see, variances in the second and fifth sub-bands are tinier and longer, which indicates that these two sub-bands are the best choices. Furthermore, our algorithm for sub-bands selection has successfully selected these two sub-bands.

Table 3 compares the classification accuracy of all subjects using DSLVQ for weighting the selected features and without using DSLVQ for weighting the features. The results show that weighting the selected features with DSLVQ method is effective in the improvement of the classification accuracy for the selected features.

CONCLUSION

In this paper, a new selection algorithm has been developed to find the optimal sub-bands. Moreover, for the performance enhancement, DSLVQ has been adopted to weigh the selected sub-bands based on their discriminative information.

It has been shown that the suggested method gives good results and outperforms the basic CSP method. As a starting point for future studies, our presented method has also demonstrated great potential for production of even better results through the application of other possible combinations of frequency sub-bands for the selection of optimal bands (e.g., overlapping sub-bands, unequal frequency band division, etc.).

ACKNOWLEDGMENT

REFERENCES

1. Liu G, Huang G, Meng J, Zhu X. A frequency-weighted method combined with common spatial patterns for electroencephalogram classification in brain-computer interface. Biomed Signal Process Control 2010;5:174-80.

2. Liao X, Yao D, Wu D, Li C. Combining spatial filters for the classification of single-trial EEG in a finger movement task. IEEE Trans Biomed Eng 2007;54:821-31.

3. Novi Q, Guan C, Dat TH, Xue P. Sub-Band Common Spatial Pattern (SBCSP) for Brain-Computer Interface. IEEE EMBS Conference on Neural Engineering; May, 2007.

4. Lemm S, Blankertz B, Curio G, Müller KR. Spatio-spectral filters for improving the classification of single trial EEG. IEEE Trans Biomed Eng 2005;52:1541-8.

5. Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Müller KR. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE Trans Biomed Eng 2006;53:2274-81.

6. Ang KK, Chin ZY, Zhang H, Guan C. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. IEEE; 2008.

7. Ang KK, Chin ZY, Zhang H, Guan C. Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. Pattern Recognit 2012;45:2137-44.

8. Thomas KP, Guan C, Lau CT, Vinod AP, Ang KK. A new discriminative common spatial pattern method for motor imagery brain-computer interfaces. IEEE Trans Biomed Eng 2009;56:2730-3.

9. Sun G, Hu J, Wu G. A Novel Frequency Band Selection Method for Common Spatial Pattern in Motor Imagery Based Brain Computer Interface. IEEE; 2010.

10. Tomioka R, Dornhege G, Nolte G, Blankertz B, Aihara K, Muller KR. Spectrally Weighted Common Spatial Pattern Algorithm for Single Trial EEG Classification. Mathematical Engineering Technical Reports; 2006.

11. Wu W, Gao X, Hong B, Gao S. Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). IEEE Trans Biomed Eng 2008;55:1733-43.

12. Akinci B. Realization of a Cue Based Motor Imagery Brain Computer Interface with its Potential Application to a Wheelchair. A Thesis Submitted to the Graduate School of Natural and Applied Sciences of Middle East Technical University; September, 2010.

13. Zhang L, Liu G, Wu Y. Wavelet and Common Spatial Pattern for EEG Signal Feature Extraction and Classification. IEEE CMCE Conference; 2010.

14. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University. July, 2003. Available from: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. [Last accessed on 2010 Apr 15].

BIOGRAPHIES

Fatemeh Jamaloo received the B.S. Degree in Electronic Engineering from Shariaty University, Tehran, Iran, in 2006, and the M.Sc. Degree in Biomedical Engineering from Shahed University, Tehran, Iran, in 2012. Since 2013, she has been as Ph.D. student at Shahed university. Her current research interests include Biomedical signal processing, BCI systems and computational neuroscience.

E-mail: fatemeh_jamaloo@yahoo.com

Mohammad Mikaeili received the B.S. Degree in Electronic Engineering from Tehran University, Tehran, Iran, in 1989, and the M.S. and Ph.D. Degrees in Biomedical Engineering from the Amirkabir University, Tehran, Iran, in 1994 and 2001, respectively. Since 2001, he has been with Shahed university,where he is currently an Assistant Professor at Engineering faculty. His current research interests include Biomedical signal processing, Sleep EEG analysis and BCI systems.

E-mail: Mikaili@shahed.ac.ir