Med Biol Eng Comput (2017) 55:1809–1818
DOI 10.1007/s11517-017-1611-4

ORIGINAL ARTICLE

Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces

Enzeng Dong[1] · Changhai Li[1] · Liting Li[1] · Shengzhi Du[2] · Abdelkader Nasreddine Belkacem[3] · Chao Chen[1]

Received: 26 April 2016 / Accepted: 20 January 2017 / Published online: 25 February 2017

Abstract Pattern classification algorithm is the crucial step in developing brain–computer interface (BCI) applications. In this paper, a hierarchical support vector machine (HSVM) algorithm is proposed to address an EEG-based four-class motor imagery classification task. Wavelet packet transform is employed to decompose raw EEG signals. Thereafter, EEG signals with effective frequency subbands are grouped and reconstructed. EEG feature vectors are extracted from the reconstructed EEG signals with one versus the rest common spatial patterns (OVR-CSP) and one versus one common spatial patterns (OVO-CSP). Then, a two-layer HSVM algorithm is designed for the classification of these EEG feature vectors, where "OVO" classifiers are used in the first layer and "OVR" in the second layer. A public dataset (BCI Competition IV-II-a)is employed to validate the proposed method. Fivefold cross-validation results demonstrate that the average accuracy of classification in the first layer and the second layer is 67.5 ± 17.7% and 60.3 ± 14.7%, respectively. The average accuracy of the classification is 64.4 ± 16.7% overall. These results show that the proposed method is effective for four-class motor imagery classification.

✉ Chao Chen
cccovb@hotmail.com

1 Key Laboratory of Complex System Control Theory and Application, Tianjin University of Technology, Tianjin 300384, China

2 Department of Mechanical Engineering, Tshwane University of Technology, Pretoria 0001, South Africa

3 Endowed Research Department of Clinical Neuroengineering, Global Center for Medical Engineering and Informatics, Osaka University, Suita 565-0871, Japan

Keywords Electroencephalography (EEG) · Motor imagery · Common spatial pattern · Hierarchical support vector machine (HSVM)

1 Introduction

Some of the people with neurological disease suffer from troubles in walking, speaking, and writing because they lose fine motor control. People with these kinds of diseases such as amyotrophic lateral sclerosis (ALS), locked-in syndrome, Lou Gehrig's disease, and high spinal cord injury lack control of their voluntary muscles. Thus, they are unable to do even simple tasks by themselves. Therefore, they cannot communicate with the environment and sometimes they are excluded from society because they are considered heavy burden [22, 24, 25]. Brain–computer interface (BCI) technology can be incorporated into medical treatments on those patients and enhance their quality of life. The mind intention of handicapped people can be detected when he performs actual or imagined movement by analyzing his/her brain signals only [9, 24] and converted into commands for sending messages or controlling home devices, which provides a higher quality of life for both disabled users and their family. Due to the safety concern of relevant techniques, noninvasive EEG-based BCI is widely used toward these assistive purposes, such as forward word spellers [16, 19], wheelchair control [7], and video games [21]. In addition, noninvasive BCIs may be useful for evaluating brain activity of severely paralyzed patients to predict the efficacy of invasive brain–machine interface [6].

EEG signals with several physiological mechanisms, such as motor imagery (MI) [2, 18], steady-state visual-evoked potential (SSVEP) [14, 26], and P300 [13], have been investigated by BCI researches. Compared with

SSVEP- or P300-based BCIs, MI methods may have higher potentiality because they are independent to an external stimulus, which allows achieving asynchronous control and communication.

For MI EEG signal processing, the traditional common spatial pattern (CSP) method, which can be interpreted in both mathematics and physiology, was recognized as an effective method for feature extraction. However, the traditional CSP method is more suitable for two-class MI EEG data classification. Furthermore, it is sensitive to noise and not suitable for small training sets [8]. In order to handle the four-class classification, the traditional CSP method was extended by computing common spatial pattern for each class against all others [5]. In addition, several approaches to improve CSP methods were proposed to address the issue of selecting optimal time frequency bands for the CSP algorithm. For example, regularization terms were added as prior knowledge in regularized CSP (R-CSP) methods [11, 15]. By adding a probabilistic counterpart of CSP, the probabilistic CSP (P-CSP) infers spatial patterns by two linear Gaussian generative models which shared the basis matrix [10]. Filter bank common spatial pattern (FBCSP) method cutting a broad frequency band into small non-overlapping filters was proposed for an MI EEG-based BCI [1].However, FBCSP method is complex to compute subject-specific frequency bands.

This paper proposes a mathematical paradigm consisting of "one versus one" (OVO) and "one versus rest" (OVR) strategies to the traditional CSP for four-class motor imagery classification. Ten common spatial patterns are calculated, and their feature vectors are extracted. An HSVM algorithm is designed to coordinate with the feature vectors. The proposed algorithm is applied on real EEG data of nine human subjects to distinguish among four motor imagery tasks. The sensorimotor cortex is the region of the cerebral cortex involved in the planning, control, and execution of voluntary movements. This cortex is responsible on motor imagery tasks; therefore, it is a critical component of sending commands and receiving feedbacks to/from muscles. Figure 1 shows the different regions of the sensorimotor cortex which were considered in this study when placing electrodes.

The rest of the paper is organized as follows. In Sect. 2, experimental tests are described and all steps of the proposed algorithm are described in detail. In Sect. 3, results of motor imagery classification are presented. Advantages and disadvantages of the proposed algorithm in different scenarios are discussed in Sect. 4.

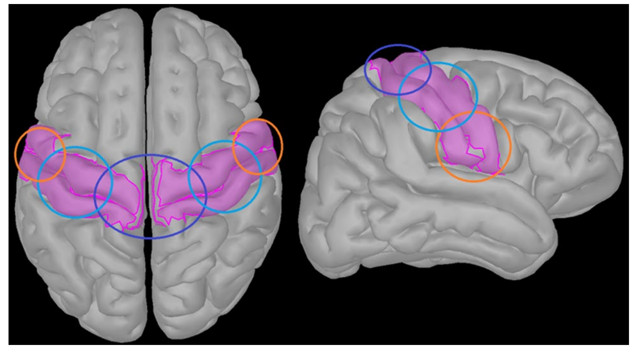


Fig. 1 Different regions of the sensorimotor cortex. The brain area highlighted in *pink* in the figure controls different body parts. The navy blue circle (top of the head) limits the region controlling feet and legs. The *pale blue circle* (center of the head) limits the region controlling hands. Face and tongue are controlled by the small orange circled region near the bottom of the cortex (located just above ears) (color figure online)

2 Method

2.1 EEG dataset

The dataset used in this study was taken from BCI competition IV-II-a [3]. It includes four motor imagery tasks: imagination of the left hand, right hand, both feet, and tongue movements of nine subjects. As shown in Fig. 2a, EEG signals were recorded from 22 Ag/AgCl electrodes and 3 monopolar electrooculogram (EOG) channels (with left mastoid serving as reference) with sampling frequency at 250 Hz, and band-pass filtered between 0.5 and 100 Hz. Power line interference was filtered by an additional 50-Hz notch filter. Timing scheme of the paradigm is shown in Fig. 2b. More detailed information about the EEG experiment can be found in [3].

The EEG data are comprised of two sessions which were recorded on different days to take into consideration the non-stationary nature of EEG data. Each session has 6 runs separated by short breaks. There are 48 trials (12 per possible class) in each run. Thus, each session is composed of 288 trials in total. For data analysis, each trial was separated and extracted by its category of motor imagery task. Then, 72 valid trials for each task were achieved. Then, 72 available trials for each task were achieved. Fivefold cross-validation was applied to counteract over-fitting. In fivefold cross-validation, the original sample (72 trials) is randomly partitioned into five subsamples. Of the five subsamples, four subsamples are used as training data; the remaining single subsample (14 trials) is retained as the validation data for testing the model. For four imaginary tasks,

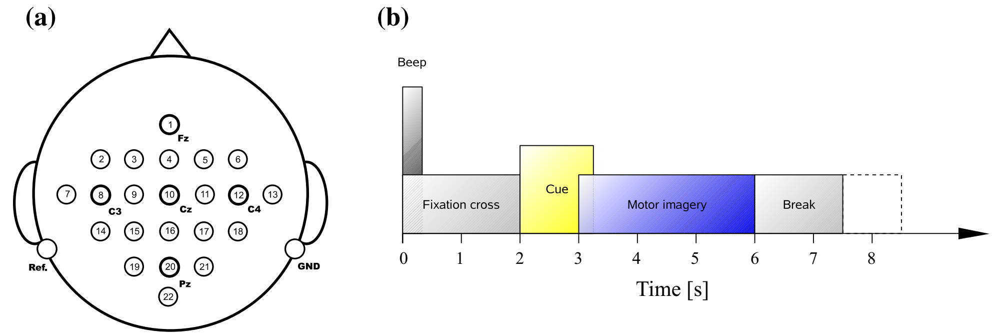


Fig. 2 Experimental paradigms. **a** Electrode positions; **b** timing scheme of the BCI paradigm

56 trials were included in test dataset. The cross-validation process is then repeated five times, with each of the five subsamples used exactly once as the validation data.

2.2 Preprocess

Motor imagery could cause event-related desynchronization (ERD) [5] and the ipsilateral hemisphere event-related synchronization (ERS) in the contra lateral hemisphere (see Fig. 1). Therefore, the μ rhythms (8–12 Hz) and β rhythms (14–30 Hz) of EEG signals in the related cortical are as would increase or decrease their amplitude and spectrum power. In addition, high-frequency component in EEG signals was usually nebulous, so the raw EEG signals were filtered by band-pass filter (3–34 Hz). Five-level wavelet package decomposition was applied to analyze the filtered EEG signals (3–34 Hz). The fifth level decomposed components $\left\{U_5^0, U_5^1, U_5^2, \ldots, U_5^{29}, U_5^{30}, U_5^{31}\right\}$ correspond to specific frequency bands. For instance, when the filtered EEG signal was considered only in the band of [3, 34] Hz, the following frequency bands are considered $\left\{\left[3, 3+\frac{f_{\text{in}}}{2^5}\right], \left[3+\frac{f_{\text{in}}}{2^5}, 3+\frac{2f_{\text{in}}}{2^5}\right], \cdots, \left[3+\frac{(2^5-1)f_{\text{in}}}{2^5}, 3+f_{\text{in}}\right]\right\}$ where f_{in} is 32 Hz. As shown in Fig. 3, the amplitudes of average reconstructed EEG signals in the band of [25, 34] Hz were very low and almost no change was recorded during motor imagery. Thus, only the frequency bands falling in [3, 24] Hz were selected and used in feature selection.

2.3 Common spatial patterns with OVO and OVR strategies

Common spatial pattern (CSP) [12] was proposed, for two-class classification for EEG-based BCIs. For the four-class classification problem considered in this study, OVO and OVR strategies were applied to adjust CSP for feature selection.

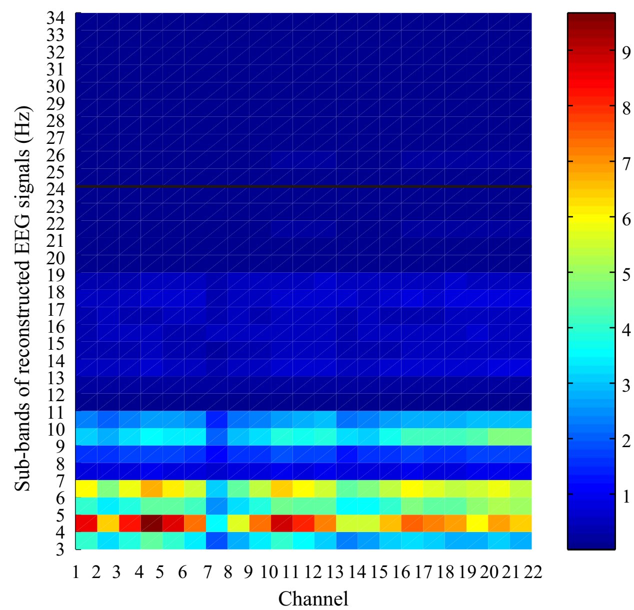


Fig. 3 Amplitudes (μV) of reconstructed EEG signals

First of all, we labeled imagination of the left hand, right hand, both feet, and tongue as classes 1, 2, 3, and 4, respectively. As shown in Fig. 4a, OVO strategy selects any two classes to form a pair to apply the traditional CSP method. By this way, a four-class classification problem is transformed into a six two-class classification problem.

Let X_i with $i \in \{1, 2\}$ denote the reconstructed EEG signal of class i. The dimension of X_i is $T \times N$ in each trial, where N and T denote the number of channels and the number of samples in time series for each channel, respectively. Note that the number of samples can be variable with respect to different subjects.

The covariance of one trial for class I is

$$C_i = \frac{X_i X_i^{\mathrm{T}}}{trace(X_i X_i^{\mathrm{T}})} \quad (1)$$

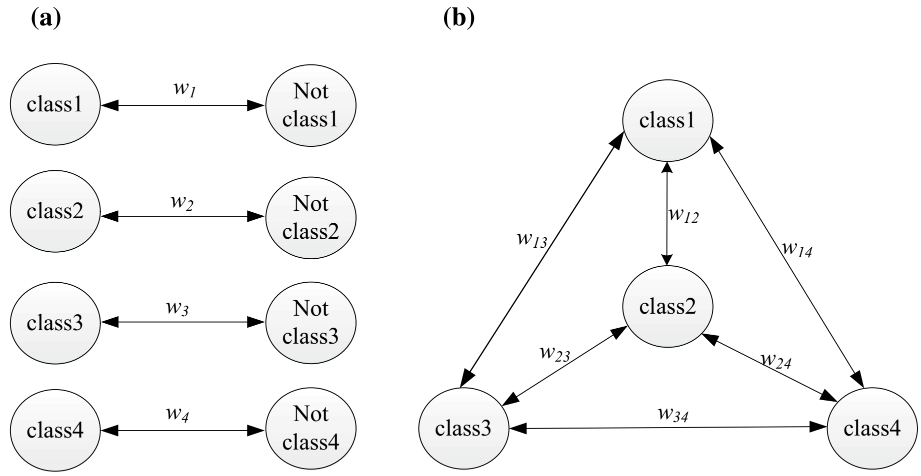


Fig. 4 Common spatial pattern strategies combined in this study: **a** OVO strategy; **b** OVR strategy

where X_i^T denotes the transpose of X_i and the trace is defined to be the sum of the elements on the main diagonal of a matrix. The spatial covariance C_i should be calculated by averaging over all trials of each group.

The composite spatial covariance C is:

$$C = C_1 + C_2 \tag{2}$$

Then, C can be factored as

$$C = U_0 \wedge U_0^{\mathrm{T}} \tag{3}$$

where U_0 is the matrix consisting of eigenvectors, and$\wedge$ is the diagonal matrix of eigenvalues. $\wedge$ is defined so that the eigenvalues were sorted in descending order.

The whitening transformation is

$$P = \wedge^{-1/2} U_0^{\mathrm{T}} \tag{4}$$

Then, C_1 and C_2 are whitened as

$$S_1 = PC_1P^{\mathrm{T}} \tag{5}$$

$$S_2 = PC_2P^{\mathrm{T}} \tag{6}$$

S_1 and S_2 share common eigenvectors. Then, S_1 and S_2 can be factored as

$$S_1 = B \wedge_1 B^{\mathrm{T}} \tag{7}$$

$$S_2 = B \wedge_2 B^{\mathrm{T}} \tag{8}$$

Then, the sum of $\wedge_1$ and $\wedge_2$ would be identity matrix

$$\wedge_1 + \wedge_2 = I \tag{9}$$

which means the largest eigenvalue in S_1 corresponded to the smallest eigenvalue in S_2, because the sum of them keeps constant 1. The eigenvectors in B will be used for classification of the two classes. The optimal feature vectors would be given for discriminating two populations of EEG when whitened EEG signals are projected to the first and the last eigenvectors. The projection matrix is

$$W_{12} = B^{\mathrm{T}} P \tag{10}$$

The projection (mapping) of a trial is given as

$$Z_{12} = W_{12} X \tag{11}$$

The rows of W_{12} can be considered as EEG source distribution vectors and the columns of W_{12} are the common spatial patterns. By decomposing the whitened EEG signals according to Eq. (11), the features for classification could be achieved. In this way, for each class of the imagined movement, only the variances of a small set of signals are needed for the classifier training.

In OVR strategy (Fig. 4b), one class was defined as target class, and the other three classes were combined as the opposite class. The covariance of each trial and whitening transformation were calculated as the traditional CSP does. However, the common spatial pattern was computed differently as

$$C = C_1 + C_{1r} \tag{12}$$

$$\text{where } C_{1r} = C_2 + C_3 + C_4 \tag{13}$$

Through whitening transformation, C_{1r} was transformed to S_{1r} and could be described as

$$S_{1r} = P_{1r} C_{1r} P_{1r}^{\mathrm{T}} \tag{14}$$

In the phase of diagonalization, S_{1r} can be factored as

$$S_{1r} = B_{1r} \wedge_{1r} B_{1r}^{\mathrm{T}} \tag{15}$$

Then, the sum of $\wedge_1$ and $\wedge_{1r}$ is

$$\wedge_1 + \wedge_{1r} = I \tag{16}$$

The projection matrix can be deduced as

$$W_1 = B_{1r}^{\mathrm{T}} P_{1r} \tag{17}$$

The mapping of a trial is

$$Z_1 = W_1 X \tag{18}$$

The feature vectors f_i could be computed as follows:

$$f_i = \log\left(\frac{VAR_i}{\sum_{i=1} VAR_i}\right) \tag{19}$$

where VAR_i denotes the variance matrix of the best projection of EEG signal Z_1, which is computed by common spatial filters. For OVO and OVR strategies, combined feature vectors are finally defined as

$$f = [f_{12}, f_{13}, f_{14}, f_{23}, f_{24}, f_{34}, f_1, f_2, f_3, f_4] \tag{20}$$

where f_{12}, f_{13}, f_{14}, f_{23}, f_{24}, f_{34} denote the EEG feature vectors corresponding to OVO strategy, which are used to train OVO classifiers, while f_1, f_2, f_3, f_4 denote the EEG feature vectors for OVR classifiers corresponding to OVR strategy.

2.4 Hierarchical support vector machine

In this study, the traditional C-support vector machine (SVM) approach was used for supervisory classification [4, 23]. The basic idea of SVM is to map the input x onto a high-dimensional feature space ($z = \phi(x)$) and look for the optimal decision hyperplane, which separates the data points into different classes with a maximum margin.

The decision hyperplane was defined as

$$w \cdot z - b = 0 \tag{21}$$

where w is the normal vector and b is the bias of the separation hyperplane. The decision hyperplane can be found by solving the following optimizing problem.

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\zeta_i \\ &s.t\ y_i[w \cdot z_i - b] \geq 1 - \zeta_i, (\zeta_i \geq 0), \quad i = 1, 2, 3, \ldots l \end{aligned} \tag{22}$$

where x_i is the i-th input sample, y_i is the class label value of x_i, l is the number of input samples, ζ_i is the slack variable that allows an example to be in the margin ($0 \leq \zeta_i \leq 1$,also called a margin error) or to be misclassified ($\zeta_i > 1$), and C is a penalty factor to be chosen by the user, a larger C corresponding to assigning a higher penalty to errors.

Eq. (21) can be solved by its dual problem using Lagrange optimization.

$$\begin{aligned} &\max \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\ &\quad s.t \sum_{i=1}^{l} y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C \end{aligned} \tag{23}$$

The solution can be calculated by

$$w = \sum_{i=1}^{Ns} y_i \alpha_i \phi(x_i) \tag{24}$$

where α_i is the Lagrange multiplier from the QP problem, Ns is the number of support vectors, and $K(x_i, x_j)$ is the kernel function.

Furthermore, radial basis kernel function, which was applied to deal with the nonlinear characteristics of EEG signal, can be described as

$$K(x_i, x_j) = \exp(-g\|x_i - x_j\|^2), \quad g > 0 \tag{25}$$

where g is the kernel parameter which denotes the gamma distribution of the transformed data. The penalty factor C controls the degree of punishment for right or wrong classification. The kernel parameter g and penalty factor C are adjusted to search for optimal separation hyperplane. Therefore, g and C play an important role in improving the correct rate and classification efficiency of the SVM. In this study, the grid search method was used to optimize g and C. To avoid over-fitting, tenfold cross-validation is used for training classifiers.

The hierarchical support vector machine paradigm is designed to optimize classification, as shown in Fig. 5. Four OVR and six OVO support vector machine classifiers are employed in the first layer and the second layer, respectively.

After preprocessing, EEG feature signals were given in input to the first-layer support vector machine which contains four OVR support vector machines. For OVR support vector machine, the classification result in OVR support vector machine maybe the "Class One" and the "Class Rest." We defined the result "Class One" as a valid classification result because the result "Class Rest" means three possible classes. Note that the valid result here does not mean this result is a correct result.

In this manner, possible results can be achieved as shown in Table 1. The possible results can be categorized into three cases:

Case 1 Only one OVR support vector machine gets valid results and other three get invalid results ("class rest").

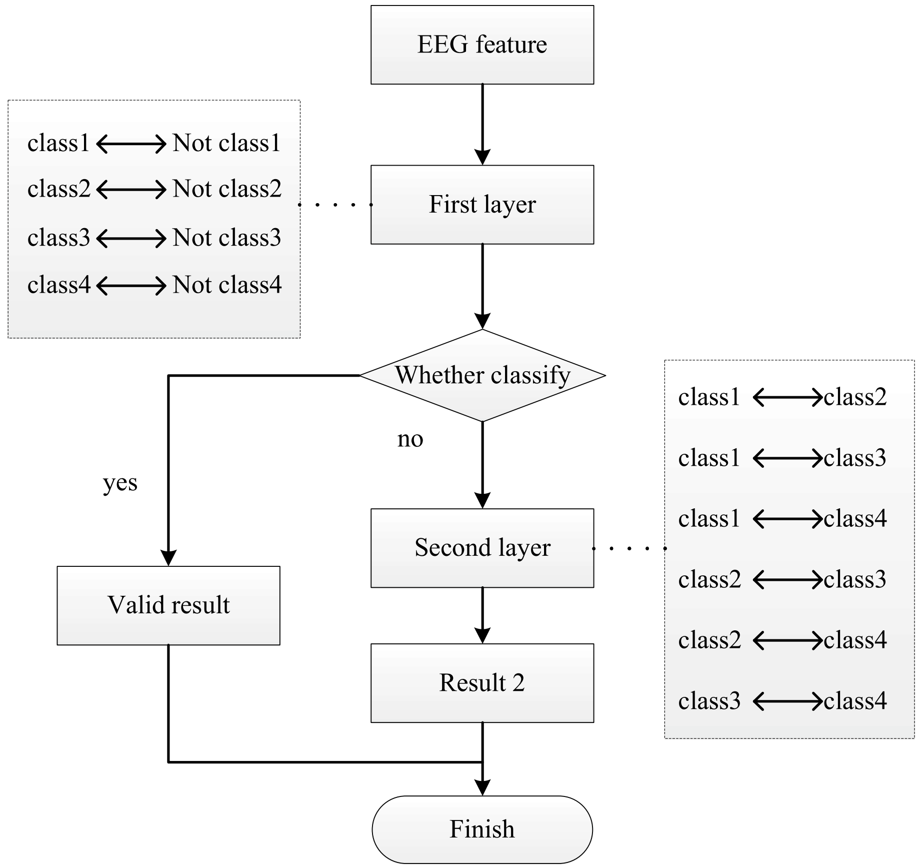
Case 2 Any two OVR support vector machines get valid result, and the other two get invalid results.

Case 3 Any other situations which are different to Case 1 and Case 2.

For Case 1, the valid result is considered as the final classification result, and the trial would be labeled. The accuracy value in first layer was calculated among these labeled trials achieved in first layer. Otherwise, the unlabeled trials are sent to the second layer. The accuracy value in second layer was calculated among the unlabeled trials achieved in first layer.

For Case 2, the EEG feature signals are entered into only one corresponding classifier according to two valid results.

Fig. 5 Flowchart of the classification process performed with the proposed HSVM algorithm



For example, Class 1 and Class 2 are the possible classes in the first layer, this trial would be sent to the classifier only for Class 1 and Class 2. The classification result is the final result and the trial is labeled.

Table 1 Possible result in first-layer classifiers

Results	Classifier			
	Left hand versus others	Right hand versus others	Both feet versus others	Tongue versus others
Case 1	1	0	0	0
	0	2	0	0
	0	0	3	0
	0	0	0	4
Case 2	1	2	0	0
	1	0	3	0
	1	0	0	4
	0	2	3	0
	0	2	0	4
	0	0	3	4
Case 3	1	2	3	0
	1	2	3	4
	1	2	0	4
	1	0	3	4
	0	2	3	4
	0	0	0	0

For Case 3, the EEG feature signals are entered into the six OVO support vector machine classifiers. The possible result is shown in Table 2. The vote rule was adopted. For situation 1, "Class one" appeared three times in OVO support vector machine classifiers. So this result was final result. Since in situation 2, "Class one" or others results just appear two times, the final result cannot be achieved. The classification of this trial was failed and counted as incorrect classification.

The final corrected rate (or fraction of correctly classified trials) was calculated as the proportion of the number of correctly labeled trials (after first- and second-layer SVM) divided by the total test number 280 (fivefold classification, 56 test trials per fold).

3 Experimental results

A test dataset containing 56 trials is considered for validating the proposed hierarchical support vector machine classifiers. The final classification results were 64.4 ± 16.7 and 69.16 ± 16.0% for sessions 1 and 2, respectively. The EEG data of sessions 1 and 2 were analyzed.

Classification results in the first layer are shown in Table 3, where the number of trials achieved valid

Table 2 Possible results of Case 3 in second-layer classifiers

Result	Classifier					
	Classifier 1_2	Classifier 1_3	Classifier 1_4	Classifier 2_3	Classifier 2_4	Classifier 3_4
Situation 1	1	1	1	3	2	3
Situation 2	1	1	4	2	2	4

Table 3 Classification results in the first layer

Subject	First layer		
	Valid result (mean ± SD)	Correct result (mean ± SD)	Accuracy (%)
S_1	14.2 ± 1.3	9.6 ± 0.9	67.6
S_2	**35.0 ± 5.3**	26.0 ± 5.1	74.3
S_3	34.2 ± 5.7	**30.2 ± 4.1**	**88.3**
S_4	21.2 ± 3.5	10.6 ± 4.1	50.0
S_5	22.6 ± 5.0	9.6 ± 2.6	42.4
S_6	20.2 ± 4.1	8.8 ± 3.7	43.6
S_7	32.4 ± 2.3	25.4 ± 5.7	78.4
S_8	33.0 ± 2.1	27.8 ± 1.9	84.2
S_9	34.0 ± 2.2	26.8 ± 2.0	78.8
Average	27.4 ± 7.8	19.4 ± 9.4	67.5 ± 17.7

Bold values represent the optimal result

Table 4 Classification results in the second layer

Subject	Second layer		
	Rest trial (mean ± SD)	Correct result (mean ± SD)	Accuracy (%)
S_1	**41.8 ± 1.3**	**28.8 ± 1.6**	68.9
S_2	21.0 ± 5.3	14.0 ± 2.9	66.7
S_3	21.8 ± 5.7	15.8 ± 5.1	72.5
S_4	34.8 ± 3.5	15.6 ± 1.1	44.8
S_5	33.4 ± 5.0	13.6 ± 2.7	40.7
S_6	35.8 ± 4.1	13.8 ± 3.3	38.5
S_7	23.6 ± 3.3	17.8 ± 3.3	**75.4**
S_8	23.0 ± 2.1	16.6 ± 2.7	72.2
S_9	22.0 ± 2.3	13.8 ± 1.3	62.7
Average	28.6 ± 7.8	16.6 ± 4.8	60.3 ± 14.7

Bold values represent the optimal result

results and correct results are 27.4 ± 7.8 and 19.4 ± 9.4 (mean ± standard deviation), respectively. The average accuracy of the first layer is 67.5 ± 17.7% in total. The largest number of valid results and correct results is 35.0 ± 5.3 for subject 2 and 30.2 ± 4.1 for subject 3, respectively. The best accuracy, 88.3%, was achieved for subject 3.

Table 4 shows the classification results in the second layer, where the “rest” results denote the trials being classified as “rest classes.” The average number of “rest” trials and correct trials is 27.4 ± 7.8 and 19.4 ± 9.4, respectively. The average accuracy is 67.5 ± 17.7%. In the second layer, subject 1 got 41.8 ± 1.3 “rest” trials, and 28.8 ± 1.6 correctly classified trials. The best accuracy is 75.4% for subject 7.

To calculate the total classification accuracy shown in Fig. 6, the numbers of correct results achieved in first layer (Table 3) and in second layer (Table 4) are added and divided by the total number of test dataset. The best accuracy is 82.1 ± 3.3% for subject 3. The average accuracy through the total 9 subjects is 64.4 ± 16.7%. A two-way ANOVA is then applied to analyze classification accuracy for the 9 subjects, and significant differences are observed ($F_{8,44} = 34.53$, $p = 1.30 \times 10^{-13}$). It can be seen that accuracy for subjects 4, 5, and 6 is lower than for the other subjects. There is no significant difference between subject 2, subject 3, subject 7, subject 8, and subject 9.

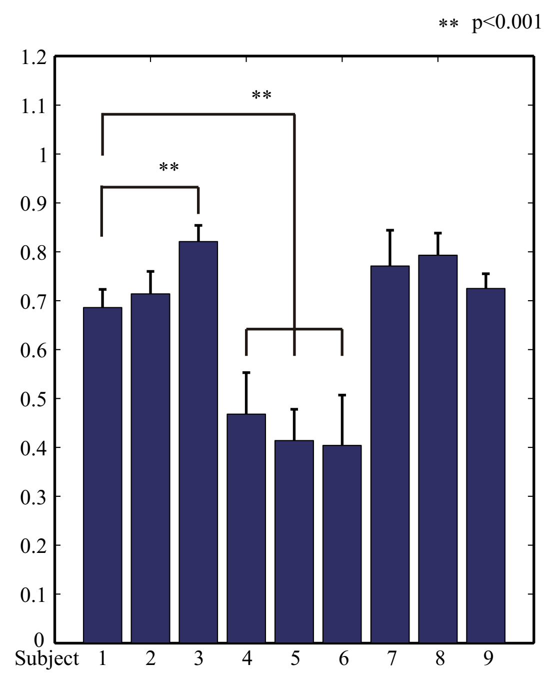


Fig. 6 Final classification accuracy achieved by the proposed approach

The classification results obtained in this study are compared with the literature [4, 10]. The final accuracy 64.4 ± 16.7% obtained in this paper for the worst session

Table 5 Comparison between our proposed algorithm and DAG SVM method

	Processing time in training phase (s)	Processing time in test phase (s)	Classification rate (%)
Proposed method	10.2	0.0001	64.4 ± 16.7 (session 1)
			69.16 ± 16.0 (session 2)
DAG SVM	5.2	0.00005	62.0 ± 15.9 (session 1)
			67.06 ± 15.4 (session 2)

(session 1) is however higher than 61.9 ± 17.7% (standard OVR-CSP method) and 62.6 ± 18.7% (filter bank method).

4 Discussion and conclusions

In this paper, two common spatial pattern strategies and hierarchical support vector machine method were proposed to process four-class motor imagery data. EEG signals were preprocessed, and the features were extracted through 10 common spatial patterns (four OVR-CSPs and six OVO-CSPs). Then, these EEG features were given in input to the hierarchical support vector machines.

Table 5 compares the performance of the proposed method with the directed acyclic graph (DAG) SVM method. Computations were carried out on a Lenovo computer (CPU 3.3 GHz). It can be seen that processing time in training phase and test phase is longer than for DAG SVM. However, processing time of test phase remains short enough for real-time applications. Furthermore, the proposed method is more accurate than DAG SVM.

Classification results demonstrated that the average classification accuracy 67.5 ± 17.7% in the first layer was higher than the 60.3 ± 14.7% accuracy achieved in the second layer. The classification process implemented in the proposed method is divided into two layers. One trial can be labeled in the first layer or in the second layer. The number of labeled results in the first OVR SVM layer reveals larger differences between one class and the other three classes in EEG signals. The number of labeled results in the first layer also correlated with the average accuracy in the first layer (correlation coefficient 0.73) and final results (correlation coefficient 0.67). Higher classification accuracy in first layer is the reason why proposed method is better than traditional SVM methods, like DAG SVM method.

The average achieved for the 9 subjects was 64.4 ± 16.7%, better than its counterpart for the traditional OVR-CSP method and filter bank method. These results prove that the proposed method is effective for four-class EEG imagery classification problems.

Testing performance of paralyzed patients in noninvasive BCIs might be useful for evaluating their brain activity to predict the efficacy of invasive clinical brain–machine interface such as for the five subjects who in this study got an average classification accuracy higher than 70%, hence satisfying the requirement criterion for real-time binary BCI [17, 20]. In addition, the final classification result (Fig. 6) showed that classification accuracy of six subjects was about and above 70% with chance level of 25% (since there are 4 classes motor imagery, the expected agreement of each class is 1/4, i.e., 25%), which suggested the proposed method is suitable for clinical and non-clinical applications.

In the near future, we are going to use our proposed algorithm in real-time motor imagery-based BCI to demonstrate its robustness and efficiency.

Acknowledgements This work was financially supported by the National Natural Science Foundation of China (61502340, 61172185), Natural Science Foundation of Tianjin City (15JCYBJC51800), and Higher School Science and Technology Development Fund Planning Project of Tianjin City (20120829).

Compliance with ethical standards

Conflict of interests The authors declare that there is no conflict of interests regarding the publication of this paper.

References

1. Ang KK, Chin ZY, Zhang H, Guan C (2008) Filter bank common spatial pattern (FBCSP) in brain–computer interface. In: IEEE international joint conference on neural networks, Hong Kong, China, pp 2390–2397
2. Brunner C, Naeem M, Leeb R, Graimann B, Pfurtscheller G (2007) Spatial filtering and selection of optimized components in four class motor imagery data using independent components analysis. Pattern Recogn Lett 28(8):957–964
3. Brunner C, Leeb R, Muller-Putz GR, Schlogl A, Pfurtscheller G (2008) BCI competition 2008-graz data set A, Institute for Knowledge Discovery (Laboratory of Brain–Computer Interfaces), Graz University of Technology
4. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. Proc IEEE Int Conf Neural Netw 2(3):1–27
5. Dornhege G, Blankertz B, Curio G, Muller K-R (2004) Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. IEEE Trans Biomed Eng 51(6):993–1002

6. Fukuma R, Yanagisawa T, Yorifuji S, Kato R, Yokoi H (2015) Closed-loop control of a neuroprosthetic hand by magnetoencephalographic signals. PLoS ONE 10(7):e0131547
7. Galan F, Nuttin M, Lew E, Ferrez PW, Vanacker G, Philips J, Millan JDR (2008) A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. Clin Neurophysiol 119(9):2159–2169
8. Grosse-Wentrup M, Liefhold C, Gramann K, Buss M (2009) Beamforming in non-invasive brain–computer interfaces. IEEE Trans Biomed Eng 56(4):1209–1219
9. Hadjidimitriou SK, Hadjileontiadis LJ (2012) Toward an EEG-based recognition of music liking using time-frequency analysis. IEEE Trans Biomed Eng 59(12):3498–3510
10. Kang H, Choi S (2014) Bayesian common spatial patterns for multi-subject EEG classification. Neural Netw 57(9):39–50
11. Kang H, Nam Y, Choi S (2009) Composite common spatial pattern for subject to subject transfer. IEEE Signal Process Lett 16(8):683–686
12. Koles ZJ (1991) The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroencephalogr Clin Neurophysiol 79(6):440–447
13. Krusienski DJ, Sellers EW, McFarland DJ, Vaughan TM, Wolpaw JR (2008) Towardenhanced p300 speller performance. J Neurosci Methods 167(1):15–21
14. Lin Z, Zhang C, Wu W, Gao X (2007) Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans Biomed Eng 54(6):1172–1176
15. Lotte F, Guan C (2011) Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. IEEE Trans Biomed Eng 58(2):355–362
16. Martens S, Leiva J (2010) A generative model approach for decoding in the visual event-related potential-based brain–computer interface speller. J Neural Eng 7(2):1393–1402
17. Pfurtscheller G, Neuper C, Birbaumer N (2005) Human brain-computer interface. In: Vaadia E, Riehle A (eds) Motor cortex in voluntary movements: a distributed system for distributed functions, Methods and New Frontiers in Neuroscience. CRC Press, Boca Raton, pp 367–401
18. Ramoser H, Muller-Gerking J, Pfurtscheller G (2010) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8(4):441–446
19. Salvaris M, Sepulveda F (2009) Visual modifications on the p300 speller BCI paradigm. J Neural Eng 6(4):046011
20. Suk HI, Lee SW (2011) Subject and class specific frequency bands selection for multiclass motor imagery classification. Int J Imaging Syst Technol 21(2):123–130
21. Tangermann M, Krauledat M, Grzeska K, Sagebaum M, Vidaurre C, Blankertz B (2008) Playing pinball with non-invasive BCI. Adv Neural Inf Process Syst 21:1641–1648
22. Thulasidas M, Guan C, Wu J (2006) Robust classification of EEG signal for brain–computer interface. IEEE Trans Neural Syst Rehabil Eng 14(1):24–29
23. Vapnik VN (2000) The nature of statistical learning theory. Springer, Berlin
24. Wolpaw JR, McFarland DJ (2004) Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc Natl Acad Sci USA 101(51):17849–17854
25. Wolpaw JR, Mcfarland DJ, Neat GW, Forneris CA (1991) An EEG-based brain-computer interface for cursor control. Electroencephalogr Clin Neurophysiol 78(3):252–259
26. Zhang D, Huang B, Li S, Wu W (2015) An idle-state detection algorithm for SSVEP-based brain-computer interfaces using a maximum evoked response spatial filter. Int J Neural Syst 25(7):1550030

Enzeng Dong received the Bachelor and Master degrees from Army Officer Academy of PLA, China, in 2000 and 2003, respectively. He received the Dr. Eng. degree from Nankai University, Tianjin, China, in 2006. From 2001, he is an Associate professor in School of Electrical Engineering, Tianjin University of Technology. He is mainly working on chaos system, brain–machine interface, and image processing.

Changhai Li is a graduate student in Electrical Engineering, Tianjin University of Technology. He is mainly working on EEG signal processing algorithm.

Liting Li is a graduate student in Electrical Engineering, Tianjin University of Technology. She is mainly working on biomedical signal processing.

Shengzhi Du graduated from Nankai University, Tianjin, China, in 2005, and received Doctor of Engineering degree. He joined Tshwane University of Technology, South Africa, as postdoctoral fellow in 2008. From 2010 to 2012, he was an Associate Professor in University of South Africa. He became a Full Professor in University of South Africa in 2012. He is mainly involved in the research fields of noninvasive brain–computer interface (BCI), image processing and pattern recognition, and computing intelligence.

Abdelkader Nasreddine Belkacem was born in Relizane, Algeria, on October 31, 1984. He received the Engineering degree in Electronics in 2006 from the University of Abdelhamid Ibn Badis, Mostaganem, Algeria, and Magister degree in Vision and Pattern Recognition in 2010 from the University of Sciences and Technology of Oran, Algeria, where he became interested in brain signals processing. In 2015, he got his PhD degree from Tokyo Institute of Technology, Japan. Currently, he is a Specially Appointed Researcher at Osaka University, Japan. His research is mainly devoted to eye-movement-based communication systems, biosignal processing, brain–machine interface, and neuroscience.

Chao Chen received the Bachelor degree in Automation, and Master degree in Pattern Recognition and Intelligent System from Northeastern University, Shenyang, China, in 2005 and 2008. In 2014, he got his PhD degree from Tokyo Institute of Technology, Japan. Currently, he is an Associate professor in School of Electrical Engineering, Tianjin University of Technology. He mainly works on invasive and noninvasive brain–computer interface (BCI), pattern recognition, and intelligent algorithm.