Classification of lower limb motor imagery based on iterative EEG source localization and feature fusion

Xiaobo Peng¹ · Junhong Liu¹ · Ying Huang² · Yanhao Mao¹ · Dong Li¹

Received: 28 July 2021 / Accepted: 10 November 2021
© The Author(s) 2021

Abstract

Motor imagery (MI) brain–computer interface (BCI) systems have broad application prospects in rehabilitation and other fields. However, to achieve accurate and practical MI-BCI applications, there are still several critical issues, such as channel selection, electroencephalogram (EEG) feature extraction and EEG classification, needed to be better resolved. In this paper, these issues are studied for lower limb MI which is more difficult and less studied than upper limb MI. First, a novel iterative EEG source localization method is proposed for channel selection. Channels FC1, FC2, C1, C2 and Cz, instead of the commonly used traditional channel set (TCS) C3, C4 and Cz, are selected as the optimal channel set (OCS). Then, a multi-domain feature (MDF) extraction algorithm is presented to fuse single-domain features into multi-domain features. Finally, a particle swarm optimization based support vector machine (SVM) method is utilized to classify the EEG data collected by the lower limb MI experiment designed by us. The results show that the classification accuracy is 88.43%, 3.35–5.41% higher than those of using traditional SVM to classify single-domain features on the TCS, which proves that the combination of OCS and MDF can not only reduce the amount of data processing, but also retain more feature information to improve the accuracy of EEG classification.

Keywords Lower limb motor imagery · Iterative EEG source localization · Feature fusion · PSO-SVM · Brain–Computer Interface (BCI)

1 Introduction

Motor imagery (MI) brain–computer interface (BCI) is an important type of BCI, which can use imagery intents to control devices and perform rehabilitation training for patients with movement disorders. It is widely used in the field of rehabilitation and human–machine interface [1]. The source distribution of MI electroencephalogram (EEG) signals on the brain cortex and their features has corresponding relationships with motion parts and motion types of the body, which makes it possible to determine the

motion content by EEG signal analysis. The previous MI-BCI researches mainly focus on the upper limb MI. Some crucial problems like channel selection, feature extraction and classification related to the MI EEG of the lower limbs still need more in-depth work.

In BCI applications, 64 or more channels are usually used to collect EEG data while only a few of them are selected for follow-up data analysis so as to reduce data amount and improve classification accuracy by reducing over-fitting [2]. In order to obtain optimal channels, many channel selection methods have been studied. Generally, the channels in the motor area, such as C3, C4, Cz, etc., are selected for the upper limb MI EEG signal analysis [3]. Varsehi et al. [4] proposed a Granger causality channel selection (GCCS) method in which the channels were sorted by the mean GC value of each channel to other channels, and the channels with higher GC values were selected. With eight selected channels, the method resulted in 93.03% accuracy, which was increased by 3.95% in comparison with correlation-based channel selection

✉ Xiaobo Peng
PENGXB@SZU.EDU.CN

¹ Shenzhen Key Laboratory of Electromagnetic Control, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China

² Shenzhen Key Laboratory of Marine Bioresources and Ecology, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China

method. Idowu et al. [5] selected optimal channels by a modified particle swarm optimization (MPSO) which was added a disturbing term to avoid local optimal solution and premature convergence. The method outperformed previous version of PSO by reducing the error rate by 10.4%. Fauzi et al. [6] used L2-norm to calculate the energy of each channel and selected the channels with high energies as the active channels for EEG classification. For the hand MI EEG dataset, C3-Cp3-C4-Cp4-Cz, C3-Cp3-C4-Cz-Cpz and Cp3-Cp4-Cpz were selected. Compared with the method without energy extraction, the proposed method improved the accuracy by up to 50%. Feng et al. [7] believed that brain areas associated with MI were not exactly the same in different frequency bands and thus proposed a common spatial pattern-rank channel selection method for multifrequency band EEG (CSP-R-MF), with different sub-frequency-bands having different selected channels. The average classification accuracy was improved by about 7% when using the proposed method compared to using CSP-R. Qi et al. [8] proposed a spatiotemporal-filtering-based channel selection (STECS) method, in which the spatiotemporal filter optimization was transformed into a Rayleigh quotient maximization problem, and channel selection was achieved by adding a sparsity-promoting regularization term. Compared with other channel selection methods, the mean classification performance of STECS on the three data sets was improved by up to 10.42%, 6.13%, and 3.72% respectively. Zhang et al. [9] inserted an automatic channel selection (ACS) layer into a convolutional neural network for MI classification. By introducing the sparse regularization, the output of the ACS layer was constrained to be sparse and the channels corresponding to the nonzero coefficients were retained for MI classification. By the method, an average accuracy of 87.2% was obtained, providing a up to 23.7% improvement with respect to other channel selection approaches. Nevertheless, the research on channel selection is more focused on the upper limbs. At the same time, it is generally achieved through approximation methods and it remains unclear how large the gap is between the selected channels and the truly optimal ones [10]. In addition, compared with the need of real-time BCI applications, the number of channels acquired by most channel selection methods is still too large and needs to be further reduced. More work is needed to find the optimal channels suitable for MI-BCI, especially for the lower limb applications which have been less studied.

Feature extraction is another critical problem to ensure the EEG classification accuracy. Commonly used feature extraction methods mainly include the following: time-domain methods, spectral methods, time–frequency-domain methods and spatial domain methods. In the time domain, Balam et al. [11] proposed a single-channel EEG

based drowsiness detection (DD) model. The EEG signal was decomposed into a series of sub-bands by wavelet packet transform (WPT). Those sub-bands in the EEG rhythm frequency ranges were extracted and then applied through the inverse WPT to get back the time-domain signals of them as feature signals. The proposed model achieved 94.45% and 85.3% accuracy on two EEG datasets, respectively, outperforming other comparison models. Wang et al. [12] adopted a one dimensional-aggregation approximation (1d-AX) method, using time-domain piecewise linear regression to represent the EEG signal and passing the data to the long short-term memory (LSTM) network for classification. On the adopted EEG data of all subjects, the classification accuracies of AX-LSTM were superior to other methods except in the experiment of the left hand/tongue group. In the spectral domain, Samuel et al. [13] compared 12 spectral domain descriptors (SDDs) with 20 time-domain descriptors (TDDs), and revealed that the SDDs worked better than the TDDs, with the best TDD achieving an accuracy of 67.05% as against 87.03% for the best SDD. By applying a linear feature combination, an optimal set of combined TDDs recorded an average accuracy of 90.68% while that of the SDDs achieved an accuracy of 99.55%. Virgilio et al. [14] applied power spectral density (PSD) and discrete wavelet transform (DWT) for EEG feature extraction. Using the proposed Spiking Neuron Network Model (SNN), a classification accuracy of 74.54% was obtained, which was 0.54% higher than the traditional neural network model. In the time–frequency domain, Gao et al. [15] utilized wavelet time–frequency analysis to calculate the energy sequence of each channel. Then, by using the channels of the scalp EEG as nodes and determining the edges between nodes according to the energy difference between channels, the human brain network model was constructed. The results of brain network analysis indicated that contralateral sensorimotor area were more closely related to the MI activities. Ortiz et al. [16] decomposed the brain rhythms in the time–frequency domain using the Empirical Mode Decomposition (EMD). The averaged power variations calculated by EMD and a second order Butterworth filter were similar, indicating that EMD could be a valid tool for the analysis of EEG signals. In the spatial domain, the common spatial pattern (CSP) method is the most commonly adopted method. Mishuhina et al. [17] decomposed the EEG signal into time stages and frequency components and extracted CSP from every decomposed time–frequency cell. On three public EEG datasets, the method improved the average classification accuracy by up to 8.6% comparing to other CSP variants. Tang et al. [18] extracted optimal frequency band of each electrode and then decomposed the signals of the band into spatial patterns to describe the differences of two classes of MI. The method achieved an average classification

accuracy of 91.25% on public data set, being 3.75% and 6.25% higher than those using original CSP and autoregressive, respectively. Despite the studies above, some researchers considered that features extracted from a single domain only provided limited information while features from different domains might contain more useful information for EEG classification [19]. In [19], multi-domain features, including Hjorth, the power spectrum estimation via maximum entropy, and time–frequency energy, were extracted and then fused into low-dimensional informative features by sparse representation. The method achieved an average accuracy of over 79%, which outperformed those single-domain feature extraction methods. Khateeb et al. [20] extracted multi-domain features by combining wavelet entropy and Hjorth parameters from EEG dataset and improved the emotion classification accuracy to 65.92%, better than the single-domain methods using entropy and Hjorth features, which were 63.62% and 64.74%, respectively. To adopt feature fusion to obtain a higher classification accuracy is now an emerging trend [21].

In the classification stage, various classification approaches have been studied, such as decision tree [22], linear discriminant analysis [23], support vector machine (SVM) [23], and those based on deep learning [24]. Among them SVM is a commonly used EEG classification method which has good generalization ability and can obtain good classification results when the number of samples is small [25]. SVM can also be combined with other methods, such as deep learning [26], Bat optimization algorithm [27], or particle swarm optimization (PSO) [28] to improve its classification accuracy.

Therefore, there seems to be a strong need to select a more appropriate channel set, explore more reasonable multi-domain features and feature classification methods for the BCI applications of lower limb MI which is more difficult in EEG signal analysis and classification, to improve the EEG classification accuracy. In this paper, we design an EEG experiment of lower limb MI, and study the methods of channel selection, multi-domain feature extraction and classification of lower limb MI EEG signals, aiming to determine the optimal channel set (OCS) and multi-domain EEG features that can reduce the EEG signal processing workload and improve the accuracy of EEG classification. The main contributions of this paper are as follows:

- (1) A novel iterative EEG source localization (ISL) method is proposed to determine the OCS suitable for lower limb MI-BCI. The OCS, consisting of FC1, FC2, C1, C2 and Cz, has a better EEG classification accuracy than the commonly used traditional channel set (TCS) C3, C4 and Cz.

- (2) A multi-domain feature (MDF) extraction method of MI EEG signals is proposed. The combination of OCS and MDF can not only reduce the amount of data processing, but also retain more useful feature information to improve the accuracy of EEG classification.
- (3) A PSO-SVM is utilized, based on the OCS and the MDFs, to improve the EEG classification accuracy of lower limb MI.

The rest of the paper is organized as follows. In Sect. 2, the lower limb MI experiment is set-up. In Sect. 3, the methods of EEG data preprocessing are briefly described. In Sect. 4, the proposed ISL method is discussed in detail. In Sect. 5, the method of feature extraction and fusion is presented, and the PSO-SVM classification method is provided. The experimental results are reported and analyzed in Sect. 6. Finally, the paper concludes in Sect. 7.

2 Experiment set-up and the data collection

2.1 Experiment platform

The lower limb MI experiment system uses the EEG signal acquisition equipment produced by Brain Products GmbH, as shown in Fig. 1. Computer 1 is used to display the experimental paradigm including the playing of inducing video for MI, and computer 2 is used to collect and display EEG signals. Other hardware includes a Brain Amp amplifier, a signal processor, a 64-channel Quik-cap that complies with the International 10–20 System of Electrode Placement, and experimental consumables such as conductive paste, syringes, medical tape, and cotton swabs. The software of the system includes Scan, E-Prime, and the inducing video player developed by us.

Before experiment, inject conductive paste into each electrode of the cap, and reduce the resistance value of each electrode to less than $5K\Omega$ to ensure that all the electrodes are in good contact, which can be judged with the help of Scan.

Subjects perform MI tasks according to the content displayed on the screen of computer 1. The raw EEG signals are measured by the electrode cap, then recorded by computer 2 after being amplified and filtered by the amplifier and the signal processor, respectively. The EEG records are furtherly used for the MI EEG classification.

2.2 Subjects and experiment environment

The subjects are 20 college student volunteers, aged 22–26 and being physically and mentally healthy. Before the experiment, each subject needs to do a week of imagery

Fig. 1 Lower limb MI experiment system

training of leg lifting to improve his motor imagery ability. He also should get enough sleep the night before the experiment, and keep his scalp clean so that the conductive paste works well.

In order to reduce the noise in the EEG signals, the experiment is carried out in an isolated small room to ensure that the environment is quiet. The subject is sitting in a luxury seat so that he is comfortable during the experiment. His eyes are at the same height as the screen and about 90 cm away from the screen. During the experiment, the subject should try not to blink or roll his eyes, and avoid unnecessary actions such as swallowing and making noise.

2.3 Experiment procedure and data recruitment

The experiment records two types of visual induced MI tasks, the rest task and the motor task. A picture of a human body sitting still is used for the rest task inducing. When the subject sees the picture, he keeps his body relaxed, without any mental activity in his mind. A video of leg raising is played as the motor task induction. When the subject sees the video, he needs to imagine once in his mind the scene of himself raising his leg.

In order to ensure that sufficient EEG data are generated, and at the same time not to make the subjects too fatigued due to the long experiment time and affect the quality of the data, 110 trials are arranged for each subject. The 110 trials are divided into five experiment groups. Each group has 22 trials, including 11 rest tasks and 11 motor tasks. The total number of trials for all subjects is 2200.

Fig. 2 Procedure of a trial

The procedure of a trial is shown in Fig. 2. Each trial lasts 10 s. When the subject gets ready, the keyboard “q” key is pressed to start the trial. During the 1st second, a red “+” appears in the middle of the screen to remind the subject to concentrate. From the 1st to the 3rd second, the screen is blank. From the 3rd to the 7th second, the motor task video or the rest task picture is randomly played on the screen. The subject performs the rest task or the motor task according to the content on the screen. From the 7th to the 10th second is the rest time. During this time, the subject relaxes himself to reduce the affection of fatigue. There is a 5-min rest time between two experiment groups. During the rest time, the subject should not make any big movements, especially not shake his head.

As mentioned above, the EEG data set of each subject includes five groups, and each group includes 22 trials. The data of each trial include the prompt segment, the blank segment, the data segment (rest task or motor task), and the rest segment. All 64 channels of EEG data, including one channel of electrooculogram (EOG) data, are collected. The sampling frequency is 128HZ. The data recruitment process are as follows:

EEG data are collected by experimental group. At the beginning of each group, the experiment assistant pressed the start button on computer 2, computer 2 starts the data collection accordingly, and the subject begins to mentally prepare. After the subject is ready, the experiment assistant presses the “q” key of computer 1, and computer 1 displays the prompt sign “+” on the screen and at the same time sends a synchronization message to computer 2. After receiving the synchronization message, computer 2 makes a synchronization mark on the received EEG signal to indicate the beginning of the trial data. After that, computer 1 displays the content in the order shown in Fig. 2 until the trial is over. In one group, computer 2 continuously records EEG signals, using the synchronization signal to identify the starting point of a trial.

3 EEG data preprocessing

The EEG signals contain a lot of noise caused by many factors such as power-line interference, electromagnetic interference, head movement, eye movement, tongue movement, etc. In addition, the data quality of the EEG signals can also be decreased if the subject is fatigue or lack of concentration during the experiment. The objective of data preprocessing is to reduce the noise and improve the signal-to-noise ratio of the EEG signals. The work of data preprocessing includes: re-reference, filtering, epoch extraction, removal of bad trials and channels, blind source separation, and removal of artifacts. They are accomplished with the help of EEGLAB [29].

Re-reference is accomplished by using the average value of the EEG signals of all channels as the reference channel.

Since MI EEG signals are mainly in the α and β rhythms, the frequency range of 1–35 Hz is studied in the paper. A 1 Hz high-pass filtering, a 35 Hz low-pass filtering, and a 50 Hz notch filtering to remove the influence of power-line interference are performed.

Epoch extraction is used to select the valid segment of the EEG signal of each trial. As shown in Fig. 2, the segment from the 3rd to the 7th second is the MI segment, and the segment from the 2nd to 3rd second is used for baseline correction. Hereby, in the paper, the EEG signal segment of 5 s, from the 2nd to 7th second, is extracted as a valid segment.

Common artifacts include signals caused by eye blink, eye movement, head movement, power-line frequency interference, EMG and so on. Independent component analysis (ICA) method is often used to remove these artifacts [30]. During the MI EEG experiment designed in the paper, subjects are required to sit still and do not have large-scale limb movements. Therefore, eye blink and eye movement have the most significant impact on EEG signals among all artifacts. They are removed by the ICA method as well as manual method both provided by EEGLAB.

4 Channel selection based on the iterative source localization (ISL)

4.1 The iterative source localization

The process of finding the source area in the brain that generates the signals collected by channels on the surface is called source localization [31]. It is an inverse problem [32] and can be approximated to a linear problem. This paper uses the weighted minimum-norm estimation (WMNE) method for EEG source localization [33]. The method uses the head volume conduction model to specify

the positions of the channels on the head surface. It adopts a distributed current density (DCD) model to solve the inverse source problem, which triangulates the continuous cerebral cortex and estimates the voltage amplitude of each triangle, also known as the source, from the surface EEG recordings. The higher the amplitude, the more active the source [34] and the closer the relationship between the source and the EEG signal.

The traditional source localization does a single sourcing operation using all the surface channel data. Many factors such as noise and data quality often affect its accuracy. To improve the accuracy of the source localization, the paper proposes an ISL method. ISL repeatedly calls source localization and source mapping which selects the channels by the resultant source area until the number of channels selected is stable. Figure 3 shows the process of the ISL.

4.2 Selection of the initial channel set

Considering that, on one hand, the brain functional areas related to motor, including primary motor area, premotor area, and supplementary motor area, are located in the central area of the top of the head. On the other hand, the electrodes in the outermost circle of the Quik-cap are far away from the brain motor areas, and their signals are easily interfered by eye movement and head movement, in order to reduce the amount of data processing and the effect of low-quality channel data, we removed the data of the 24 channels on the outermost circle of the Quik-cap, as well as the EOG channel IO and those at the edge of the

Fig. 3 Flowchart of the ISL

frontal and occipital regions, namely AF3, AF4, PO3, POz, and PO4. That is to say, 34 channels in the middle area of the brain as shown in Fig. 4 are selected as the initial channel set S .

The paper uses time–frequency analysis to verify the validity of the EEG signals of S . All the 64 channel signals and the middle 34 channel signals of all subjects are averaged, respectively, to create two data sets, a 64-channel data set and a 34-channel data set. In order to improve the accuracy of the signals, the data of 1 s before the MI task is used for baseline correction to reduce the influence of signal drift. Then, short-time Fourier transform (STFT) is performed on these two data sets, and two time–frequency diagrams are obtained, as shown in Fig. 5a, b, respectively. It can be seen that the event-related desynchronization (ERD) phenomenon of the middle 34-channel signals is more obvious than that of the whole 64-channel signals, which means that the 34-channel signals are less affected by noise and more closely related to the lower limb MI.

4.3 Generation of source distribution images

After source localization, a source distribution image is obtained, on which the color of a pixel is used to indicate the current source density (mA/m^2) of the pixel as a source, that is, the source intensity. The brighter the color, the more active the source is, as shown in Fig. 6. Here, in order to show the positions of the sources on the brain, the source distribution image is displayed together with the brain template.

At different time during a trial, the distribution and the intensities of the sources of the same subject are also slightly different. In this paper, the time-domain signal

curves of all channels are drawn and the point with the largest amplitude on the curves is extracted as the source localization time, as shown in Fig. 7. At the source localization time, the EEG signal is the strongest and the source localization result is obvious. Different subjects have different source localization times. The subject shown in Fig. 7 has a source localization time of $t = 0.302$ s.

The source localization is performed for each subject at his source localization time. Figure 8 shows the 20 source distribution images of the first source localization. Here, the intensity threshold is set to 10%, so the 10% sources with the lowest intensity are ignored in each image. Because the objective of the ISL algorithm is to find the distribution of the sources with the highest intensity, removing the sources of the lowest intensity can reduce the amount of data without affecting the result of channel selection.

4.4 Extraction of the source area

The source distribution images are converted to grayscale ones. Figure 8 shows the resultant grayscale source distribution images of the first source localization. These 20 grayscale images are superimposed and averaged and the resultant source image I_a of this source localization is obtained. Figure 9 shows I_a and its gray histogram. In order to reduce the dimension of the features that need to be processed and the workload of feature classification, generally only the top n active channels are selected for follow-up EEG analysis. Since the 64 channels are roughly evenly placed on the head surface, n channels occupy approximately $s = n/64$ of the area of I_a .

To find the top n active channels, it is actually to determine a gray threshold k that meets the following condition: on I_a , the ratio of the area of all pixels whose gray values are greater than threshold k to the area of I_a is s . The paper uses Eq. (1) to calculate k and thus to determine the source distribution:

$$k = \max \left\{ g_i : P(g_i) = \frac{\sum_{m=g_i}^{255} f(m)}{\sum_{n=0}^{255} f(n)} \geq s; i, m, n \in [0, 255] \right\} \quad (1)$$

where k is the gray threshold that needs to be calculated, g_i is a gray value between $[0, 255]$, $f(m)$ is the distribution of gray value m , and $P(g_i)$ is the proportion of pixels with a gray value greater than g_i . As shown in Fig. 9, $P(g_i)$ can be obtained by integrating the gray distribution in the range of gray value not less than g_i .

Considering that the data volume of 5 channels can achieve a good balance between the amount of feature information and the workload of data processing, in this paper n is set to 5. When $n = 5$, $s = 7.81\%$. Then, $k=173$ is

Fig. 4 34 channels selected as the initial channel set

Fig. 5 Time–frequency diagrams of 64 and 34 channels

Fig. 6 Source distribution image

obtained by Eq. (1). Using $k=173$ as the threshold, I_a is inverted and binarized to obtain the source area image I_s as shown in Fig. 10.

4.5 Source mapping

Source mapping is to find the channels on the scalp above the source area. It is based on an intuitive fact that the scalp surface being closest to a source area has the most intensive EEG signals generated by the source, and the channels above the source area get the best EEG signals. This is also the reason why channels above the motor area, such as C3, C4, Cz, etc., are commonly selected for EEG signal analysis.

By superimposing I_s and the electrode placement map, the channels above I_s are obtained. As shown in Fig. 11, it

Fig. 7 Source localization time

Fig. 8 Grayscale source distribution images of the first source localization

Fig. 9 Source image and its gray histogram

Fig. 10 Source area image of the first source localization

can be seen that the channels in the source area include 7 channels: Fz, FC1, FC2, C1, C2, Cz and CP2. They are the result of the first iteration of the ISL, and at the same time they constitute the initial channel set for the second iteration.

Fig. 11 Resultant channels of the first iteration of the ISL

4.6 Determination of the OCS

Repeat the source localization with Fz, FC1, FC2, C1, C2, Cz and CP2 as the initial channel set. The result of the second iteration includes the channels FC1, FC2, C1, C2 and Cz, which are more concentrated in the middle area than the first iteration.

Figure 12 is the result of the third iteration. The resultant channels are the same as the second iteration: FC1, FC2, C1, C2 and Cz. According to the termination condition of the ISL, the iteration is terminated and the OCS selected for the lower limb MI are: FC1, FC2, C1, C2, Cz.

Fig. 12 Result of the channel selection

5 Feature extraction and classification

5.1 Feature extraction

We hope to use as little feature data as possible to express the imagery intentions as accurately as possible. In the paper, time-domain features, frequency-domain features, time–frequency features and spatial domain features are extracted and merged into multi-domain features to express the imagery intentions, which perform better than single-domain features [28].

(1) Time-domain (TD) feature extraction

As mentioned above, the valid segment of each trial is 5 s, of which the first 1 s is the baseline segment, and the subsequent MI time is 4 s. With the sampling rate of 128 Hz, there are totally 640 sampling points for each trial.

The EEG signal is 8–13 Hz band-pass filtered first, and the frequency band power is calculated to obtain the TD features. The square of the voltage of the sampling point on the TD EEG signal curve is used to characterize the power. For the EEG signal curve of a trial of a subject, the maximum power $P_{\max}$, the minimum power $P_{\min}$, and the average power P_{mean} are extracted as TD features according to Eqs. (2), (3), (4).

$$P_{\max_ij} = \max\{v_{ijk}^2, k \in [1, n]\} \quad (2)$$

$$P_{\min_ij} = \min\{v_{ijk}^2, k \in [1, n]\} \quad (3)$$

$$P_{\text{mean_ij}} = \frac{1}{n} \sum_{k=1}^n v_{ijk}^2 \quad (4)$$

where $n = 640$ is the number of sampling points on the EEG curve of a trial. $i \in [1, 5]$ is the channel number in the OCS. $j \in [1, 110]$ is the trial number. $k \in [1, n]$ is the sampling point number. v_{ijk} is the voltage of sampling point k of trial j of channel i , and thus $P_{\max_ij}$, $P_{\min_ij}$, and

$P_{\text{mean_ij}}$ are the maximum power, the minimum power, and the average power of trial j of channel i , respectively.

Three TD features are extracted for each channel in the OCS. Each subject has 15 TD features for each trial. Then a TD feature matrix $B_1(2200 \times 15)$ can be generated for all subjects and all trials.

(2) Frequency-domain (FD) feature extraction

For the FD EEG signal of each channel at 0–40 Hz, the power spectral density (PSD) is calculated using the Welch method, and then its mean value PSD_{mean} , standard deviation PSD_{std} , average power PSD_{avg} , kurtosis PSD_{kur} , and skewness PSD_{ske} are extracted as frequency-domain features. For the 5 channels of the OCS, 25 FD features are extracted. A FD feature matrix $B_2(2200 \times 25)$ is generated for all subjects and all trials.

(3) Time–frequency-domain (TFD) feature extraction

Discrete Wavelet Transform (DWT) is used to decompose the signals of the OCS, with db4 as the wavelet basis function. According to the Nyquist sampling theorem, the valid frequency is 64 Hz when the sampling rate is 128 Hz. After four-layer decomposition, five frequency bands of 0–4 Hz, 4–8 Hz, 8–16 Hz, 16–32 Hz, 32–64 Hz are obtained, as shown in Fig. 13. The detail coefficient cD3 on the third level represents the 8–16 Hz frequency band of the EEG signals, which contains the α frequency band (8–13 Hz) that has the most obvious ERS/event-related synchronization (ERS) phenomenon. Hereby, the average energy of the frequency band of cD3 is selected as the TFD feature.

One TFD feature is extracted for each channel of the OCS and a TFD feature matrix $B_3(2200 \times 5)$ is obtained for all subjects and all trials.

(4) Spatial domain (SD) feature extraction

CSP works well for two-class classification. This paper extracts the SD features of the EEG signals based on the CSP algorithm. First a spatial filtering is performed on the rest task data and the motor task data, then the variance scaling method is used to expand the distance between these two types of data and maximize the variance between

Fig. 13 DWT decomposition

them. The variance of the voltage values of 640 sampling points is output as the SD feature.

Each channel has a SD feature, and a SD feature matrix $B_4(2200 \times 5)$ can be obtained for all subjects and all trials.

(5) Feature fusion and multi-domain (MD) feature generation

Assemble the above four feature matrices by Eq. (5) to form the MD feature matrix.

$$B = [B_1, B_2, B_3, B_4] \quad (5)$$

According to the size of B_1 , B_2 , B_3 , and B_4 , it can be known that the size of the MD feature matrix B is 2200×50 . B can be written as: $B = [b_i | i = 1, \dots, 50]$, where b_i is the i th column feature vector.

For any channel i , combine its corresponding feature column vectors by Eq. (6) to realize the feature fusion and obtain the fusion feature vector t_i :

$$t_i = \left(\sum_{j=3(i-1)+1}^{3(i-1)+3} b_j + \sum_{j=5(i-1)+16}^{5(i-1)+20} b_j + b_{j+40} + b_{j+45} \right) \quad (6)$$

where $i \in [1, 5]$. t_i is the fusion feature vector corresponding to the i th channel in the OCS. b_j is the j th column feature vector in B .

After combination, the MD feature matrix $T(2200 \times 5) = [t_1, t_2, t_3, t_4, t_5]$ is obtained. Figure 14 shows the distributions of randomly sampled feature data of channel pairs. The data points are like (t_{ij}, t_{ik}) where i is the trial number and j, k are channel numbers. It can be seen that the points are obviously clustered according to whether they belong to the rest task or the motor task, which is helpful to find a specific plane to classify them.

5.2 Feature classification

As mentioned earlier, SVM is an excellent binary classifier which is also adopted in this paper. Radial basis kernel function (RBF) is often selected as the kernel function of SVM because it requires few parameters and is flexible. It has good performance regardless of sample size and has good anti-interference ability. Because of this, it is often selected as the kernel function in the application of SVM-based MI EEG signal classification and achieved excellent classification accuracy [35]. In this paper, we also use RBF as the kernel function of SVM. However, some studies have also achieved better classification performance using kernel functions other than RBF. For example, Chui et al. [36] used the cross correlation kernel $K_{\text{corr},ij}$ and the convolution kernel $K_{\text{conv},ij}$ to construct a Mercer kernel function KDDC and performed drowsiness detection of drivers based on electrocardiogram data. The accuracy of KDDC was proved better than typical kernels including linear, quadratic, third order polynomial, and Gaussian RBF by 17–63%, respectively. Therefore, customizing a kernel function suitable for the characteristics of the EEG signals to obtain a better classification performance is also our future work.

Considering the clustering distribution characteristics of the fused feature data, in this paper we adopt PSO based SVM for feature classification. PSO is used to optimize the penalty parameter c and the kernel function parameter g of SVM to improve its classification ability. In the PSO, the population size is set to be 20 and 100 iterations are performed to optimize the SVM parameters. The final optimization result is $c = 0.9221$, $g = 0.7832$.

We use tenfold cross-validation for performance validation. The multi-domain feature matrix is divided into 10 row blocks. Nine blocks are used as training data and the

Fig. 14 Clustering distribution characteristics of feature data

Table 1 Classification results for different channels and features

Feature type	Classification accuracy			
	OCS(FC1, FC2, C1, C2, Cz)		TCS(C3, C4, Cz)	
	Traditional SVM [35]	PSO-SVM	Traditional SVM [35]	PSO-SVM
TD [11]	81.59%	84.56%	81.13%	84.13%
PSD FD [37]	82.56%	85.68%	81.22%	83.25%
DWT TFD [38]	82.41%	85.08%	81.91%	81.93%
CSP SD [18]	83.87%	85.91%	83.03%	85.83%
MD	84.80%	86.57%	83.63%	85.89%

remaining one block is used as test data.

6 Results and analysis

6.1 Classification results

Table 1 shows the classification results of the method proposed in the paper comparing to other methods. Two data sets are used for comparison. One is the EEG signals on the OCS generated by the ISL, and the other one is the signals on TCS commonly used by existing works. For each data set, five kinds of features are extracted: TD features, FD features, TFD features, SD features, and MD features. Two kinds of SVM are adopted for the classification: Traditional SVM and PSO-SVM. In the traditional SVM, after many tries and adjustments, the penalty parameter c and the kernel function parameter g are set to 1.0 and 0.8, respectively. As a comparison, in PSO-SVM the parameters are optimized to $c = 0.9221$, $g = 0.7832$.

PSO-SVM is also compared to two popular ensemble methods, Bagging and Gradient Boosting. Here Random Forest and Gradient Boosting Machine are used. Each classification method is tested on five kinds of features: TD features, FD features, TFD features, SD features, and MD features. All the features are extracted from the OCS. Table 2 shows the results of these classification methods.

6.2 Result analysis

(1) Channel selection

The purpose of ISL is to find the most active source area with the highest correlation with the EEG data, so as to obtain the channel set with the most obvious EEG features. After three iterations, the stable OCS is obtained: FC1, FC2, C1, C2, and Cz. As shown in Fig. 12, compared with the TCS, C3, C4, and Cz, which are often used for the upper limb BCI applications, the distribution area of OCS is more concentrated in the center on the coronal section of the brain, which is consistent with the widely recognized lower limb motor area in the brain. This may imply that the OCS we selected are suitable for the expression of lower limb MI EEG data.

It can be seen from Table 1 that when the same classification method, traditional SVM (T-SVM) or PSO-SVM, is used, for the same domain, the classification accuracies on the OCS for TD features, FD features, TFD features, SD features and MD features are increased by 0.46%/0.43%, 1.34%/2.43%, 0.50%/3.15%, 0.84%/0.08%, and 1.17%/0.68%, respectively, comparing to those on the TCS, as shown in Fig. 15. It proves that the classification effect of the OCS is better than the TCS.

(2) Multi-domain feature extraction

From Sect. 5.1, we can see that the amount of MD features is the same as that of the TFD features and the SD features, smaller than that of the TD features

Table 2 Classification results of different classification methods

Feature type	Classification accuracy		
	Bagging [39]	Gradient Boosting [39]	PSO-SVM
TD [11]	69.16%	77.77%	84.56%
PSD FD [37]	72.03%	77.90%	85.68%
DWT TFD [38]	69.88%	75.24%	85.08%
CSP SD [18]	76.21%	76.79%	85.91%
MD	77.69%	80.15%	86.57%

Fig. 15 Improvements of classification accuracies of the OCS comparing to the TCS

and the FD features.

As shown in Table 1, using the same classification method and channel set, the classification accuracies of MD features increase by 0.93–3.21%, 0.66–2.21%, 0.60–2.50%, 0.06–3.96%, respectively, comparing to single-domain features, as shown in Fig. 16a.

As shown in Table 2, the classification accuracies of MD features using Bagging, Gradient Boosting and PSO-SVM are increased by 1.48–8.53%, 2.25–4.91%, and 0.66–2.01%, respectively, comparing to single-domain features, as shown in Fig. 16b.

The results show that using MD features, better classification accuracies than single-domain features are achieved in the case of no increase in the amount of data.

(3) Feature classification

It can be seen in Table 1 that (1) When using the same channels and the same features, the PSO-SVM method has higher classification accuracies

compared with the traditional SVM algorithm, as shown in Fig. 17a. (2) Using MD features and the OCS data, PSO-SVM achieves the highest accuracy of 86.57%, which is significantly higher than other methods in the table.

It can be seen in Table 2 that PSO-SVM gets better classification accuracies than Bagging and Gradient Boosting for the different features, as shown in Fig. 17b, and the PSO-SVM with the MD features gets the best result.

In summary, based on the OCS and MD features, the PSO-SVM achieves a satisfying classification accuracy.

7 Conclusions

In this paper, an iterative source localization method is proposed for the channel selection for the EEG signal classification of lower limb motor imagery. Based on the method, five channels FC1, FC2, C1, C2, and Cz are selected as the OCS, which perform better than the commonly used traditional channel set. For the EEG signals of the OCS, TD features, FD features, TFD features, and SD features are extracted and fused into MD features expressed as a MD feature matrix. Finally, the PSO-SVM classification method is used to classify the multi-domain fusion features. The results show that the classification accuracy is 88.43%, 3.35–5.41% higher than those of using traditional SVM to classify single-domain features on the TCS, which proves that the combination of OCS and MD

Fig. 16 Improvements of classification accuracies of MD features comparing to single-domain features

Fig. 17 Improvements of classification accuracies of PSO-SVM

features can not only reduce the amount of data processing, but also retain more feature information to improve the accuracy of EEG classification.

Future work related to lower limb MI-BCI should focus on the development of more accurate classification methods, including customizing the basis function of SVM to make it more suitable for the characteristics of lower limb MI EEG data, and expanding the classification ability from binary classification to multi-class classification. At the same time, more efficient and automated methods of data processing should be studied to realize online real-time feature recognition.

Acknowledgments The research was supported by National Natural Science Foundation of China (U1713204) and Shenzhen Basic Research Foundation (JCYJ20180305125327627).

Declarations

Conflict of interest We all declare that we have no conflict of interest in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Xu K, Huang Y, Duann J (2019) The sensitivity of single-trial mu-suppression detection for motor imagery performance as compared to motor execution and motor observation performance. *Front Hum Neurosci* 13:302
2. Baig MZ, Aslam N, Shum HPH (2020) Filtering techniques for channel selection in motor imagery EEG applications: a survey[J]. *Artif Intell Rev* 53:1207–1232
3. Shan H, Xu H, Zhu S, He B (2015) A novel channel selection method for optimal classification in different motor imagery BCI paradigms[J]. *Biomed Eng Online* 14:93
4. Varsehi H, Firoozabadi SMP (2021) An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality[J]. *Neural Netw* 133:193–206
5. Idowu OP, Adelopo O, Ilesanmi AE, Li X, Samuel OW, Fang P, Li G (2021) Neuro-evolutionary approach for optimal selection of EEG channels in motor imagery based BCI application[J]. *Biomed Signal Process Control* 68:102621
6. Fauzi H, Azzam MA, Shapii MI, Kyoso M, Khairuddin U, Komura T (2019) Energy extraction method for EEG channel selection[J]. *TELKOMNIKA* 17(5):2561–2571
7. Feng J, Jin J, Daly I, Zhou J, Niu Y, Wang X, Cichocki A (2019) An Optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system[J]. *Comput Intell Neurosci* 5:8068357
8. Qi F, Wu W, Yu Z, Gu Z, Wen Z, Yu T, Li Y (2021) Spatiotemporal-filtering-based channel selection for single-trial EEG classification[J]. *IEEE Trans Cybern* 51(2):558–567
9. Zhang H, Zhao X, Wu Z, Sun B, Li T (2021) Motor imagery recognition with automatic EEG channel selection and deep learning[J]. *J Neural Eng* 18:016004
10. Narayanan AM, Patrinos P, Bertrand A (2021) Optimal versus approximate channel selection methods for EEG decoding with application to topology-constrained neuro-sensor networks[J]. *IEEE Trans Neural Syst Rehabil Eng* 29:92–102
11. Balam VP, Chinara S (2021) Automatic classification methods for detecting drowsiness using wavelet packet transform extracted time-domain features from single-channel EEG signal. *J Neurosci Methods* 347:108927
12. Wang P, Jiang A, Liu X, Shang J, Zhang L (2018) LSTM-based EEG classification in motor imagery tasks[J]. *IEEE Trans Neural Syst Rehabil Eng* 26:2086–2095
13. Samuel OW, Geng Y, Li X, Li G (2017) Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors[J]. *J Med Syst* 41:194
14. Virgilio GCD, Sossa AJH, Antelis JM, Falcón LE (2020) Spiking neural networks applied to the classification of motor tasks in EEG signals[J]. *Neural Netw* 122:130–143
15. Gao Z, Wang Z, Ma C, Dang W, Zhang K (2018) A wavelet time-frequency representation based complex network method for characterizing brain activities underlying motor imagery signals[J]. *IEEE Access* 6:65796–65802
16. Ortiz M, Iáñez E, Contreras-Vidal JL, Azorín JM (2020) Analysis of the EEG rhythms based on the empirical mode decomposition during motor imagery when using a lower-limb exoskeleton. A case study[J]. *Front Neurobot* 14:48
17. Mishuhina V, Jiang X (2021) Complex common spatial patterns on time-frequency decomposed EEG for brain-computer interface[J]. *Pattern Recognit* 115:107918
18. Tang Z, Li C, Wu J, Liu P, Cheng S (2019) Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI. *Front Info Technol Electron Eng* 20:1087–1098
19. Xu C, Sun C, Jiang G, Chen X, He Q, Xie P (2020) Two-level multi-domain feature extraction on sparse representation for motor imagery classification. *Biomed Signal Process Control* 62:102160
20. Khateeb M, Anwar SM, Alnowami M (2021) Multi-domain feature fusion for emotion classification using DEAP dataset. *IEEE Access* 9:12134–12142
21. Chu Y, Zhao X, Zou Y, Xu W, Han J, Zhao Y (2018) A decoding scheme for incomplete motor imagery EEG with deep belief network[J]. *Front Neurosci* 12:680
22. Roy S, Rathee D, Chowdhury A, McCreadie K, Prasad G (2020) Assessing impact of channel selection on decoding of motor and cognitive imagery from MEG data[J]. *J Neural Eng* 17(5):056037
23. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, Yger F (2018) A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update[J]. *J Neural Eng* 15(3):031005
24. Tabar YR, Halici U (2017) A novel deep learning approach for classification of EEG motor imagery signals[J]. *J Neural Eng* 14(1):016003
25. Razzak I, Hameed IA, Xu G (2019) Robust sparse representation and multiclass support matrix machines for the classification of motor imagery EEG signals[J]. *IEEE J Tran Eng Health Med* 7:1–8

26. Li Y, Li J, Pan J (2019) Hyperspectral image recognition using SVM combined deep learning. *J Internet Technol* 20(3):851–859
27. Selim S, Tantawi MM, Shedeed HA, Badr A (2018) A CSP AM-BA-SVM approach for motor imagery BCI system[J]. *IEEE Access* 6:49192–49208
28. Ma Y, Ding X, She Q, Luo Z, Potter T, Zhang Y (2016) Classification of motor imagery EEG signals with support vector machines and particle swarm optimization[J]. *Comput Math Methods Med* 4941235:1–8
29. Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis[J]. *J Neurosci Methods* 134:9–21
30. Islam MK, Rastegarnia A, Yang Z (2016) Methods for artifact detection and removal from scalp EEG: a review[J]. *Neurophysiol Clinique/Clin Neurophysiol* 46(4–5):287–305
31. Asadzadeh S, Rezaii TY, Beheshti S, Delpak A, Meshgini S (2020) A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities[J]. *J Neurosci Methods* 339:108740
32. Cui S, Duan L, Gong B, Qiao Y, Xu F, Chen J, Wang C (2019) EEG source localization using spatio-temporal neural network[J]. *China Commun* 16(7):131–143
33. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python[J]. *Front Neurosci* 7:267
34. Ojeda A, Kreutz-Delgado K, Mullen T (2018) Fast and robust block-sparse bayesian learning for EEG source imaging[J]. *Neuroimage* 174:449–462
35. Bousseta R, Tayeb S, Ouakouak EI, Gharbi M, Regragui F, Himmi MM (2018) EEG efficient classification of imagined right and left hand movement using RBF kernel SVM and the joint CWT_PCA. *AI & Soc* 33:621–629
36. Chui KT, Tsang KF, Chi HR, Ling BWK, Wu CK (2016) An Accurate ECG-Based transportation safety drowsiness detection scheme[J]. *IEEE Trans Industr Inf* 12(4):1438–1452
37. Kim C, Sun J, Liu D, Wang Q, Paek S (2018) An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI. *Med Biol Eng Compu* 56(9):1645–1658
38. Ji N, Ma L, Dong H, Zhang X (2019) EEG signals feature extraction based on DWT and EMD combined with approximate entropy [J]. *Brain Sci* 9:201
39. Gonzalez S, Garcia S, Ser DJ, Rokach L, Herrera F (2020) A practical tutorial on bagging and boosting based ensembles for machine learning: algorithms, software tools, performance study, practical perspectives and opportunities. *Info Fusion* 64:205–237

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.