Biomedical Signal Processing and Control 71 (2022) 103247

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Joint spatial and temporal features extraction for multi-classification of motor imagery EEG

Xueyu Jia [1], Yonghao Song [1], Lie Yang, Longhan Xie [*]

Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China

ARTICLE INFO

Keywords:
Brain-computer interface (BCI)
Electroencephalogram (EEG)
Motor imagery (MI)
Spatial filtering
Convolutional neural network (CNN)

ABSTRACT

The application of brain-computer interface (BCI) has always been limited by low decoding accuracy due to excessive noise in electroencephalogram (EEG) signals. The traditional methods employ some representative features while losing too much information. Deep learning methods have achieved good results, but subject to the insufficient ability of extracting discriminative features from EEG. In this paper, we propose a novel decoding framework that effectively uses spatial and temporal information by time-contained spatial filtering and spatial–temporal analysis network (TSF-STAN) for EEG multi-classification tasks. Firstly, the TSF with the joint one-versus-rest (Joint-OVR) strategy is given to map the signal to a new space, where each category can be more easily distinguished, while preserving the time-domain characteristics. Next, the STAN is designed to extract discriminative spatial and temporal features further with convolutional layers, and then perform classification. Detailed experiments have been carried out to verify the effectiveness of our framework on BCI competition IV-2a and IV-2b datasets of motor imagery (MI) EEG. The results show that our method has outperformed recent outstanding algorithms, with the average accuracy of 83.0% on IV-2a and 88.0% on IV-2b. Spatial and temporal information is well used to obtain better performance, which has a good potential of EEG decoding for application of BCI.

1. Introduction

In recent years, brain-computer interface (BCI) becomes one of the most advanced neurophysiological technologies that can connect brains with peripheral devices [1,2]. Numerous attempts have been made to explore the practice of BCI, among which motor imagery (MI) has attracted wide attention because the motor cortex responds when a person mentally simulates a given movement of limbs without muscle activation [3]. Repetitive training based on MI promotes neural remodeling, even after severe nerve damage [4,5]. Combined with some external assistive equipment, the MI-based system has great potential to help people who suffer from stroke, spinal cord injury, and amyotrophic lateral sclerosis (ALS) improve their life quality [5,6]. The key to ensuring the practicality and robustness of BCI systems is the accurate decoding of user intent. Specifically, feature extraction and classification are the two most concerning goals. Some researchers employed event-related desynchronization and event-related synchronization (ERD/ERS) to classify mental states [7,8], because the sensorimotor cortex produces the attenuation of oscillatory brain activity within particular frequency bands [9,10], which is easily distinguished when the subject imagines moving different sides of the body. However, the detection of the ERD/ERS pattern is easily affected by the low signal-to-noise ratio (SNR) of EEG itself [11].

Subsequently, many researchers began to focus on the study of feature extraction methods such as fast Fourier transform (FFT) and wavelet transform [12]. This type of method innovatively uses EEG frequency domain information to facilitate MI classification. There are also a few methods such as common spatial pattern (CSP) use spatial enhancement to increase the difference of different categories [13]. Several extensions based on CSP further confirm the significance of spatial information for EEG decoding. Ang *et al.* [14] proposed the filter bank common spatial pattern (FBCSP) implemented CSP on different frequency bands and selected the ones more relevant for classification. Besides, Guo *et al.* [15] presented filter band component regularized

* Corresponding author.
E-mail addresses: xueyujia_scut@outlook.com (X. Jia), eeyhsong@mail.scut.edu.cn (Y. Song), 201810100415@mail.scut.edu.cn (L. Yang), xielonghan@gmail.com (L. Xie).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bspc.2021.103247
Received 22 July 2021; Received in revised form 17 September 2021; Accepted 6 October 2021
Available online 20 October 2021

common spatial patterns (FCCSP) to address the problems of the poor robustness of CSP with small training samples and obvious performance variability between frequency bands for different individuals. These feature extraction methods have achieved amazing results with some classifiers, such as linear discriminant analysis (LDA) [16], support vector machine (SVM) [17], Bayesian classifiers [18], multi-layer perceptron (MLP) [19]. However, considerable information is lost in the process of focusing only on specific features. Another problem is that methods similar to spatial enhancement only deal with binary classification. Although traditional one-versus-rest (OVR) splits the multi-classification task, it still cannot be collaboratively optimized.

Recently, deep learning attracts more and more attention for its excellent performance in computer vision and speech recognition [20,21]. People also try to employ a convolutional neural network (CNN) to obtain deep representation for EEG classification. Schirrmeister *et al.* [22] proposed deep ConvNets to learn the causal contributions of features in different frequency bands to decoding decisions. Moreover, Amin *et al.* [23] designed a multi-layer CNN (MCNN) fusing CNNs with different characteristics and architectures to extract different types of features representing the EEG data at various abstract levels. However, the ability to extract discriminative features of CNN is limited because the structure of convolution is easily affected by the low SNR and non-stationary of EEG. Compared with the pure CNNs, the design using prior knowledge, that is, the distinguishing feature, seems to have good potential for EEG decoding. Sakhavi *et al.* [24] developed a new classification framework called C2CM for MI-based BCI by introducing envelope representation of EEG using Hilbert transformation and passing it through a CNN. Additionally, Kim *et al.* [25] presented a new form of input for the CNN, which constructed magnitude and phase-based features with Continuous Wavelet Transform (CWT). Although these methods are impressive, the decoding is not accurate enough for the practice of BCI.

Therefore, in order to make better use of the inherent characteristics of EEG and the advantages of the deep model, we propose a method called time-contained spatial filtering and spatial–temporal analysis network (TSF-STAN) for the multi-classification of EEG. The TSF maps raw EEG data into a new space and obtain a spatial–temporal representation, then the STAN extracts discriminative spatial and temporal features further from the previous representation. To address multi-classification, we propose the joint one-versus-rest (Joint-OVR) strategy, with which a multi-classification task is converted into multiple binary classification tasks. Then the TSF generates the same number of mapping spaces based on these tasks and all the mapping spaces are merged into one space to transform the raw EEG. In this way, each category is more distinguishable while preserving temporal information. The classification results of our method outperform the other outstanding algorithms on BCI competition IV-2a MI data set [26] and IV-2b MI data set [27], which proves that our method has robust performance to deal with unstable EEG.

The significant contributions of this paper are threefold.

1) We propose time-contained spatial filtering (TSF) that increases the inter-category difference of EEG, with the temporal features well preserved.
2) We present a CNN-based spatial–temporal analysis network (STAN) to utilize discriminative spatial and temporal features further and classify different categories of EEG in an end-to-end process.
3) A Joint-OVR strategy is designed with TSF to transform the original signal into a spatial–temporal representation, which is suitable for optimizing multi-classification tasks.

The rest of the paper is organized as follows. Section 2 describes the overall architecture and details of the proposed method. Section 3 presents the experiment setting and the results of the evaluation. Discussion and analysis are given in Section 4, and Section 5 concludes the paper.

2. Methodology

In this section, we divide the method into two parts, including the TSF with the Joint-OVR, and the STAN. The algorithm of the TSF with the Joint-OVR is explained in detail firstly. Next, we describe the architecture and principles of the STAN. The proposed method is finally summarized.

2.1. Time-contained spatial filtering with joint one-versus-rest

The spatial correlations between EEG channels are essential features that influence classification accuracy. In this part, we propose the time-contained spatial filtering (TSF) to construct a mapping space, in which each category of EEG is exceptionally distinguished. At the same time, the temporal information of EEG is still retained. However, it is limited to only handle binary classification tasks for spatial filtering. To address this issue, the joint one-versus-rest (Joint-OVR) is proposed and applied to the TSF for multi-classification. OVR means that one category of EEG is considered one class, and the remaining categories of EEG are viewed as another class. In other words, multi-classification is separated into multiple binary classifications. For the Joint-OVR, it not only divides the multi-classification into multiple bi-classifications, but also merges the spatial filters of multiple bi-classifications into one. The procedures of the TSF with the Joint-OVR for N-classification are as follows, shown in Fig. 1(a).

1) The EEG from all recorded channels is filtered out noise using a bandpass filter with a bandwidth of 4–40 Hz. The data after noise removal is regarded as the raw data $E \in \mathbb{R}^{B\times C\times T}$ for the input of the method, where B is all the trials for training, C is the number of the recorded EEG channels, and T is the number of the single-channel sampling points.
2) Because of the Joint-OVR, the multi-classification is divided into N bi-classifications. There are N identical copies of the raw data for the N bi-classifications,

$$E_i = E, (i = 1, 2, \cdots, N) \tag{1}$$

Besides, for each bi-classification, the data E_i is divided into two classes E_{i1} and E_{i2} without overlap,

$$E_{i1} \cup E_{i2} = E_i, E_{i1} \cap E_{i2} = \varnothing \tag{2}$$

3) We calculate the covariance matrix calculation function of two-class of every bi-classification,

$$c_{ij} = \frac{E_{ij}\cdot E_{ij}^T}{trace(E_{ij}\cdot E_{ij}^T)}, (i = 1, 2, \cdots, N; j = 1, 2) \tag{3}$$

where $trace(E)$ represents the trace of matrix E and $C_{ij} \in \mathbb{R}^{B\times C\times C}$ represents the expectation of covariance matrix of E_{ij}. Average all of C_{ij} to obtain $\overline{C_{ij}} \in \mathbb{R}^{C\times C}$,

$$C_{ic} = \overline{C_{i1}} + \overline{C_{i2}}, (i = 1, 2, \cdots, N) \tag{4}$$

where C_{ic} represents sum of spatial covariance matrix of $\overline{C_{i1}}$ and $\overline{C_{i2}}$.
4) Since the mixed space covariance matrix C_{ic} is a positive definite matrix, the eigendecomposition is performed by the singular value decomposition theorem,

$$C_{ic} = U_{ic}\Lambda_{ic}U_{ic}^T, (i = 1, 2, \cdots, N) \tag{5}$$

where U_{ic} is the eigenvector matrix, Λ_{ic} represents the diagonal matrix of eigenvalues, and the eigenvalues are arranged in descending order. Whitening conversion P_i is obtained by

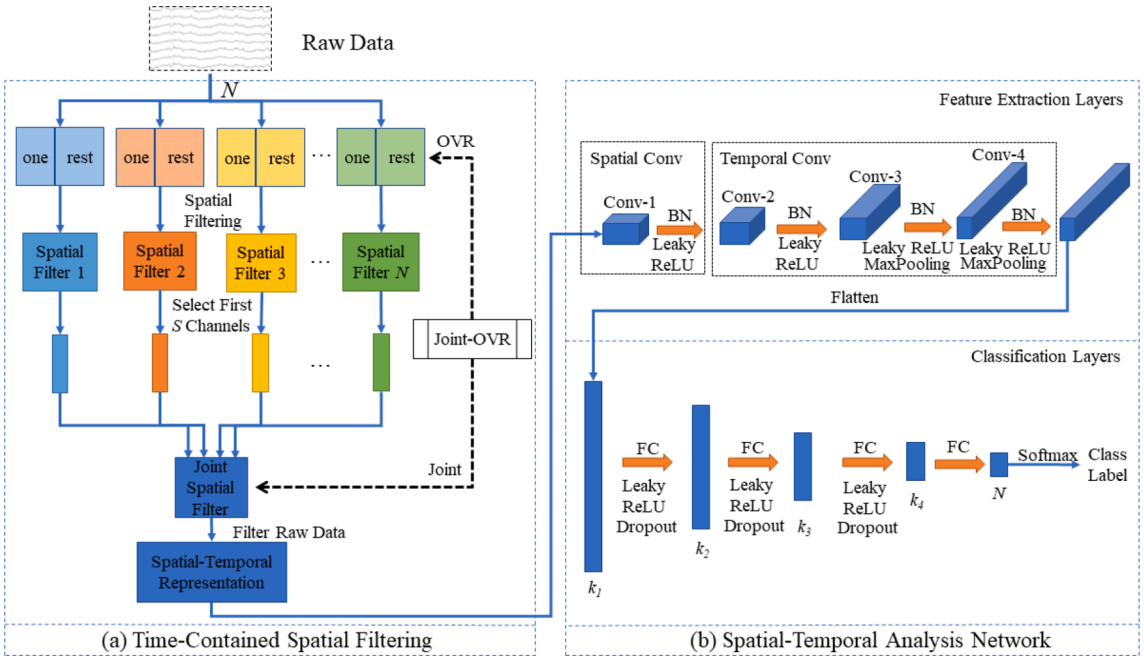

Fig. 1. The Architecture diagram of the proposed method for N-classification.

$$P_i = \frac{1}{\sqrt{\Lambda_{ic}}} U_{ic}^T, (i = 1, 2, \cdots N) \tag{6}$$

Apply the matrix P_i to $\overline{C_{i1}}$ and $\overline{C_{i2}}$, we get

$$S_{i1} = P_i \overline{C_{i1}} P_i^T, S_{i2} = P_i \overline{C_{i2}} P_i^T, (i = 1, 2, \cdots, N) \tag{7}$$

S_{i1} can be obtained by eigen decomposition,

$$S_{i1} = B_i \Lambda_{i1} B_i^T, (i = 1, 2, \cdots, N) \tag{8}$$

where B_i is the eigenvectors and Λ_{i1} is the eigenvalues of S_{i1}. Due to orthogonality,

$$B_i B_i^T = I \tag{9}$$

where I is an identity matrix. We can prove that S_{i1} and S_{i2} share the same eigenvectors, and the sum of two corresponding eigenvalues is always equal to I, shown as follows.

$$\begin{aligned} I &= B_i^T \sqrt{\Lambda_{ic}^{-1}} U_{ic}^T U_{ic} \Lambda_{ic} U_{ic}^T \left(\sqrt{\Lambda_{ic}^{-1}} U_{ic} \right)^T B_i \\ &= B_i^T P_i C_{ic} P_i^T B_i \\ &= B_i^T P_i C_{i1} P_i^T B_i + B_i^T P_i C_{i2} P_i^T B_i \\ &= B_i^T S_{i1} B_i + B_i^T S_{i2} B_i \\ &= \Lambda_{i1} + B_i^T S_{i2} B_i, (i = 1, 2, \cdots, N) \end{aligned} \tag{10}$$

where $B_i^T S_{i2} B_i$ is also a diagonal matrix with the value of $I - \Lambda_{i1}$ and we represent it as Λ_{i2}. B_i diagonalize S_{i2} so B_i is also the eigenvectors of S_{i2}.

$$B_i^T S_{i2} B_i = \Lambda_{i2}, (i = 1, 2, \cdots, N) \tag{11}$$

The left side of the equal sign is multiplied by B_i, and the right side is multiplied by B_i^T,

$$S_{i2} = B_i \Lambda_{i2} B_i^T, (i = 1, 2, \cdots, N) \tag{12}$$

where Λ_{i2} is the eigenvalues of S_{i2}. The sum of Λ_{i1} and Λ_{i2} is always equal to I. When the value of Λ_{i1} gets larger, the value of Λ_{i2} becomes smaller. Therefore, the features of one class and the other class are significantly distinguished.

5) For the eigenvector matrix B_i, when S_{i1} has the largest eigenvalue, S_{i2} has the smallest eigenvalue. Therefore, the matrix B_i can be used to classify the binary task, and the mapping matrix can be obtained by

$$W_i = B_i^T P_i, (i = 1, 2, \cdots, N) \tag{13}$$

where the mapping matrix $W_i \in \mathbb{R}^{C \times C}$ is the corresponding spatial filter. The channels of the filter are usually selected in some studies [28]. The first S channels and the last S channels of the mapping matrix correspond to the largest S eigenvalues and the smallest S eigenvalues, respectively. They have similar distinguishability, so we choose the first S channels, which also reduces the calculation cost, and form a new matrix $W_i' \in \mathbb{R}^{S \times C}$.

6) According to the Joint-OVR, all the spatial sub-filters of bi-classifications are merged into one. Therefore, all of the mapping matrixes W_i' are concatenated into one mapping matrix W,

$$W = \left[(W_1')^T, (W_2')^T, \ldots, (W_N')^T \right]^T \tag{14}$$

where $W \in \mathbb{R}^{(N*S) \times C}$ is the two-dimensional joint feature spatial filter. The symbol (*) is different from ($\times$), (*) means multiplication of values, and ($\times$) distinguishes different dimensions.

We multiply the EEG raw data and the mapping matrix W to obtain the filtered data.

$$Z = WE \tag{15}$$

where $Z \in \mathbb{R}^{B \times (N*S) \times T}$ is the intermediate spatial–temporal representation and the input of the following STAN.

2.2. Spatial-temporal analysis network

The architecture of the spatial–temporal analysis network (STAN) is

shown in Fig. 1(b). It consists of feature extraction layers and classification layers. The spatial–temporal representation is extracted with temporal and spatial features through the feature extraction layers. Then the classification layers can use these effective features to output the class label. The design details of STAN are given as follows.

1) ***Kernel of The First Layer:*** The size and stride of the kernel of the first layer, which is also designed as the spatial convolutional layer, should be related to the selected channel number S of the mapping matrix. Otherwise, different mapping spaces interfere with each other during the convolution process, and it is not helpful for feature extraction and classification. The size and stride of the kernel are equal to 1 in the time dimension and S in the channel dimension. The detail of the relation of the kernel and the spatial–temporal representation is shown in Fig. 2.
2) ***Feature Extraction layers:*** There are 4 layers in the feature extraction layers. Each layer has a convolutional layer, a batch normalization layer (BN), and a leaky rectified linear unit layer (Leaky ReLU). In addition, the third and fourth layers also have a max-pooling layer. The first layer is designed as the spatial convolutional layer, named for the only convolution in the channel dimension, to explore spatial features of the spatial–temporal representation. The following three layers are regarded as temporal convolutional layers to extract temporal features, and they only convolve in the time dimension. All network parameters of the feature extraction layer are shown in Table 1. Finally, the data is flattened into a 1-D feature vector with the shape of k_1 and fed into the classification layers.
3) ***Classification layers:*** The classification layers, consisting of several fully connected layers (FC), Leaky ReLUs, and Dropout layers, can decode the data and output the classification result. The vector shapes before and after the fully connected layer roughly satisfy the following relationship,

$$k_i = 4k_{i+1}, (i = 1, 2, 3) \tag{16}$$

where k_i is the shape of data. The flattened data from feature extraction layers is decoded through the classification layers and predicted the label after the softmax layers.

In summary, the Joint-OVR separates multi-classification into multiply binary classifications. The TSF constructs spatial filters for each bi-classification. All of the spatial filters are concatenated into one according to the Joint-OVR, and the raw data is mapped into a space, in which each category of EEG is distinguished and temporal information is still retained. The filtered data is the spatial–temporal representation. It is processed by the STAN to further extract the temporal features and spatial features and decoded into an intuitive result of intention. The code is available at https://github.com/Jia-Xueyu/TSF-STAN.

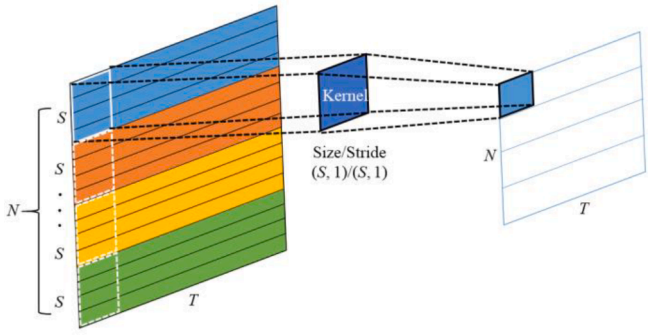

Fig. 2. The relation diagram of the spatial–temporal representation and the kernel of the first convolutional layers: The raw number of the size and stride of the kernel is the same as the selected channel number of single binary classification.

Table 1
Some parameters of STAN.

No. of layer	Sublayer	Output	Kernel	Stride
Layer 1(Spatial Conv)	Conv	16	(S,1)	(S,1)
	BN + Leaky ReLU		–	
Layer 2(Temporal Conv-1)	Conv	32	(1,23)	(1,3)
	BN + Leaky ReLU		–	
Layer 3(Temporal Conv-2)	Conv	64	(1,17)	(1,1)
	BN + Leaky ReLU		–	
	Max-pooling	–	(1,6)	(1,6)
Layer 4(Temporal Conv-3)	Conv	128	(1,7)	(1,1)
	BN + Leaky ReLU		–	
	Max-pooling	–	(1,2)	(1,2)
	(All Leaky ReLU parameters are 0.2, all Dropout parameters are 0.3)			

3. Experiments

3.1. Experimental dataset

3.1.1. BCI competition IV-2a 4-class MI dataset

The BCI IV 2a data set [26] contains four MI task classes (left hand, right hand, foot, and tongue) from nine subjects. The data for each subject consists of two sessions, one for training and one for testing. Each session has 288 trials, with an average of 72 trials per class. All of the data was recorded at a sampling rate of 250 Hz using 22 Ag/AgCl electrodes, and the amplifier sensitivity was set to 100 μV.

3.1.2. BCI competition IV-2b 2-class MI dataset

This data set [27] includes two classes (motor imagery of left hand and right hand) EEG data from 9 subjects. For each subject, 5 sessions were provided, whereby the first two sessions contained training data without feedback, and the last three sessions were recorded with feedback. Each session has 120 trials in the first two sessions and 80 trials in the last three sessions. The data were recorded from 3 EEG channels (channel C3, Cz, and C4) with a sampling frequency of 250 Hz.

3.2. Parameters setting

In all the next experiments, T was set to 1000, which meant 1000 sampling points. The data of BCI IV 2a per trial from 2 s to 6 s was intercepted as a sample for the input of the method. As for the BCI IV 2b data set, we used the truncated data from 3 s to 7 s as a sample for the first two sessions and from 3.5 s to 7.5 s for the last three sessions.

Besides, S was set to 4 for BCI IV 2a data set and 2 for BCI IV 2b data set. The value of S was determined according to the actual test results for different data sets. N was 4 for BCI IV 2a and 2 for BCI IV 2b.

3.3. Performance metrics

We use accuracy and the Cohen's Kappa coefficient [29] as the criterion for evaluating experimental performance. Kappa coefficient is used for consistency tests and can also be used to measure classification accuracy. The calculation of the Cohen's Kappa coefficient k is based on confusion matrix and defined as follows,

$$k = \frac{p_o - p_e}{1 - p_e} \tag{17}$$

where p_o is the sum of the number of samples correctly classified in each category divided by the total number of samples, which is the overall classification accuracy. p_e stands for chance coincidence rate.

Besides, we use Student's t-test (t-test) [30] hypothesis testing to compare the significant differences between the two classification result samples. The statistic t can be obtained by

$$t = \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad t(m+n-2) \tag{18}$$

$$S_w = \frac{1}{m+n+1}\left[(m-1)S_1^2 + (n-1)S_2^2\right] \tag{19}$$

where $\bar{x}$ and $\bar{y}$ are the mean of the first and another result samples, respectively, m and n are the numbers of two samples, S_1^2 and S_2^2 are the variances, and $t(m+n-2)$ indicates that the t statistic obeys the t distribution with $m+n-2$ degrees of freedom. By querying the t-quantile table, the probability value called $p-value$ can be obtained. When the $p-value$ is larger than 0.01 and less than 0.05, there is a significant difference in the classification results of the two groups. $p-value < 0.01$, the difference is very significant.

3.4. Results

3.4.1. TSF increases the inter-category difference of EEG

In order to prove that TSF increased the inter-category difference of EEG, which is beneficial to STAN to extract discriminative features better and classify, we compared the results of STAN and TSF + STAN on BCI IV 2a data set, as shown in Table 2. The average classification accuracy of TSF + STAN is 0.17 ($p-value < 0.01$) higher than that of STAN. Besides, the statistical results of accuracy are expressed and normalized in the form of confusion matrixes, as shown in Fig. 3. It can be seen that the confusion matrix of TSF + STAN has a larger value on the diagonal than that of STAN and a smaller value on the off-diagonal corner. At every position on the diagonal, TSF + STAN has a higher value than STAN. The performance of both TSF + STAN and STAN in predicting the feet is the worst among the four types of labels. STAN performs the best when predicting the right hand, while TSF + STAN performs the best when predicting the left hand and tongue. From the results and the confusion matrixes, we can see that the existence of the TSF significantly improves the accuracy of classification. In brief, the TSF maps the raw data to a space to distinguish categories so that it boosts the decoding and classification of STAN.

3.4.2. TSF retains temporal features

The TSF is based on CSP, while the first and last few channels of filters are usually used with the log-variance transformation [28] after spatial projection. In order to ensure that the number of selected channels of the mapping matrix is the same in CSP and TSF, the first two channels and the last two channels of spatial filters are selected to obtain $W^{''}{}_i \in \mathbb{R}^{4\times C}$ from Equation (13) in CSP. The feature vector is calculated by the following equation,

$$v_i = log\frac{diag\left(W^{''}{}_i E_i (W^{''}{}_i E_i)^T\right)}{tr\left[W^{''}{}_i E_i (W^{''}{}_i E_i)^T\right]}, (i = 1,2,3,4) \tag{20}$$

where $v_i \in \mathbb{R}^{B\times 4}$ is the feature vector after CSP, which loses the temporal information. $diag(.)$ returns the diagonal elements of the square matrix; $tr[\cdot]$ returns the sum of the diagonal elements in the square matrix. The feature vectors are concatenated into one feature matrix according to Joint-OVR,

$$M = [v_1, v_2, v_3, v_4]^T \tag{21}$$

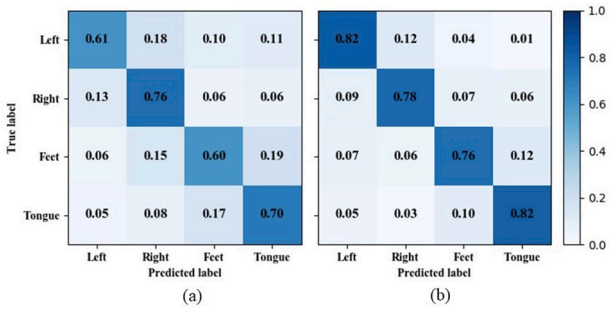

Fig. 3. Confusion matrix of the accuracy results of STAN and TSF + STAN on BCI IV 2a data set. (a) STAN; (b) TSF + STAN.

where $M \in \mathbb{R}^{B\times 4\times 4}$ is the feature matrix used for multi-classification.

The shape of the spatial–temporal representation $Z \in \mathbb{R}^{B\times 16\times 1000}$ is different from the shape of the feature matrix $M \in \mathbb{R}^{B\times 4\times 4}$. From the comparison of Z and M, we can see that the length of the feature matrix M in the time dimension is far less than the spatial–temporal representation Z. In fact, M has no time dimension. Its second dimension represents the feature vectors. Compared with CSP, TSF keeps temporal information of EEG while both are utilizing spatial filtering to increase the inter-category difference. To prove it, we tried our best to design a CNN similar to STAN, which was called easy-STAN temporarily, to compare the effects of TSF and CSP. The structure of the easy-STAN is shown in Fig. 4. The feature matrix M and the spatial–temporal representation Z based on the BCI IV 2a data set are fed into the easy-STAN. The results of the comparison are shown in Table 3.

From Table 3, we can see that the result of each subject has superior for the TSF. The average classification accuracy of the TSF is 0.099 higher than that of CSP ($p-value < 0.01$). Although the structure of the easy-STAN network may not be suitable for the data processed by the TSF, it can be seen that the data processed by the TSF has a higher classification accuracy for CNN. Based on the classification of the same network, the result of the TSF is higher than that of CSP, indicating that the TSF contains more features that can be extracted by CNN than CSP. Combining the viewpoints mentioned before, the spatial–temporal representation Z maintains the time-series information. In contrast, the feature matrix M only contains 4 feature vectors after Joint-OVR, which loses the information of the time dimension of the raw data. In other words, compared with CSP, the TSF retains the temporal information of EEG while distinguishing the EEG category.

3.4.3. STAN extracts discriminative features

In order to verify that the STAN extracts more discriminative features, we use our method, which is TSF + STAN, to compare the effect with some traditional non-neural network classifiers such as SVM and two classic CNNs (LeNet-5 [31] and VGG11 [32]). We compare the results of TSF + STAN, TSF + SVM, TSF + LeNet-5, and TSF + VGG11 on the BCI IV 2a data set, as shown in Fig. 5.

From Fig. 5, based on the TSF, the classification accuracy of SVM, LeNet-5 and VGG11 is far inferior to the STAN—The average accuracy of SVM, LeNet-5, VGG11 is 0.4 ($p-value < 0.01$), 0.269 ($p-value < 0.01$) and 0.117 ($p-value < 0.01$) lower than that of the STAN, respectively.

Table 2
The results of STAN and TSF-STAN on the BCI IV 2a data set.

Accuracy (Kappa)	S1	S2	S3	S4	S5	S6	S7	S8	S9	
STAN	0.783 (0.711)	0.450 (0.261)	0.817 (0.787)	0.550 (0.388)	0.433 (0.234)	0.567 (0.392)	0.683 (0.572)	0.750 (0.663)	0.883 (0.843)	0.657 (0.539)
TSF-STAN	**0.883 (0.841)**	**0.817 (0.748)**	**0.922 (0.866)**	**0.776 (0.660)**	**0.633 (0.542)**	**0.675 (0.552)**	**0.900 (0.864)**	**0.950 (0.924)**	**0.917 (0.888)**	**0.830 (0.765)**

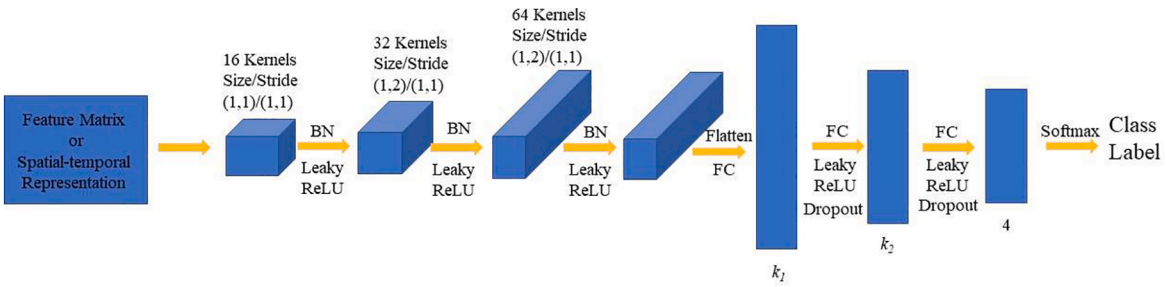

Fig. 4. The structure of the easy-STAN.

Table 3
The results of the traditional CSP and the TSF on the BCI IV 2a data set.

Accuracy (Kappa)	S1	S2	S3	S4	S5	S6	S7	S8	S9	
CSP	0.717 (0.613)	0.517 (0.344)	0.817 (0.753)	0.533 (0.380)	0.467 (0.389)	0.517 (0.361)	0.800 (0.723)	0.733 (0.641)	0.650 (0.530)	0.639 (0.526)
TSF	**0.725 (0.654)**	**0.767 (0.695)**	**0.821 (0.797)**	**0.687 (0.602)**	**0.523 (0.517)**	**0.550 (0.337)**	**0.865 (0.813)**	**0.873 (0.812)**	**0.832 (0.763)**	**0.738 (0.666)**

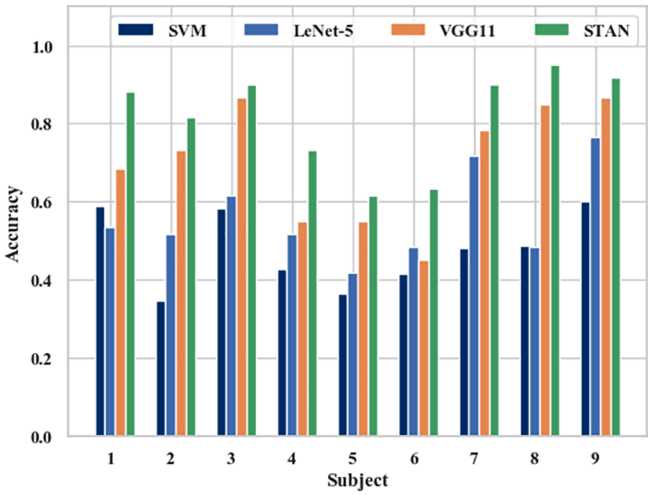

Fig. 5. The Classification accuracy of some classifiers and the STAN based on the TSF on the BCI IV 2a data set.

Compared with some non-neural network classifiers, the STAN is more able to explore more information of EEG due to its unique convolution structure so that it can use more spatial and temporal information to classify accurately. In comparison with LeNet-5 and VGG11, we can find that the STAN has a higher accuracy because STAN extracts more discriminative features than LeNet-5 and VGG11. In a word, the STAN extracts more discriminative features from the spatial–temporal representation to improve the decoding accuracy.

3.4.4. *Ablation on the spatial–temporal analysis network*

The STAN further explores the spatial and temporal information of the spatial–temporal representation. Every sublayer of the feature extraction layers is beneficial to the improvement of the results. We design ablation experiments to prove that each sublayer is indispensable. For the convenience of representation, each of our sub-layers is represented in order as Spatial Conv, Temporal Conv-1, Temporal Conv-2, and Temporal Conv-3, shown in Table 1. Therefore, the corresponding methods that lack this sub-layer are called Non-Spatial Conv, Non-Temporal Conv-1, Non-Temporal Conv-2, and Non-Temporal Conv-3. These methods are evaluated on the BCI IV 2a data set, and the result is shown in Table 4.

From Table 4, we can see that Non-Spatial Conv, Non-Temporal Conv-1, Non-Temporal Conv-2, Non-Temporal Conv-3 have average classification accuracy of 0.787, 0.740, 0.735, 0.765, respectively, while the complete method has an average accuracy of 0.830 ($p-value < 0.01$). Missing any convolutional sublayer would reduce the final classification accuracy. Each sub-layer of the STAN is functional and indispensable. The effect of feature extraction of every convolutional layer is different on the improvement of the result. They all extract features, and improve the classification accuracy as a whole.

Table 4
The results of the traditional CSP and the TSF on the BCI IV 2a data set.

Accuracy (Kappa)	S1	S2	S3	S4	S5	S6	S7	S8	S9	
Non-Spatial Conv	0.850 (0.793)	0.800 (0.726)	0.873 (0.811)	0.700 **(0.662)**	0.613 (0.501)	0.583 (0.369)	0.863 (0.733)	0.900 (0.859)	0.900 (0.866)	0.787 (0.702)
Non-Temporal Conv-1	0.800 (0.725)	0.783 (0.700)	0.817 (0.754)	0.733 (0.639)	0.517 (0.355)	0.500 (0.304)	0.867 (0.815)	0.789 (0.762)	0.850 (0.798)	0.740 (0.650)
Non-Temporal Conv-2	0.800 (0.728)	0.733 (0.636)	0.800 (0.735)	0.650 (0.565)	0.550 (0.384)	0.533 (0.339)	0.883 (0.844)	0.867 (0.814)	0.800 (0.730)	0.735 (0.642)
Non-Temporal Conv-3	0.783 (0.710)	0.700 (0.589)	0.900 **(0.866)**	0.650 (0.566)	0.600 (0.450)	0.583 (0.397)	0.850 (0.797)	0.933 (0.906)	0.883 (0.842)	0.765 (0.680)
Our method	**0.883 (0.841)**	**0.817 (0.748)**	**0.922 (0.866)**	**0.776** (0.660)	**0.633 (0.542)**	**0.675 (0.552)**	**0.900 (0.864)**	**0.950 (0.924)**	**0.917 (0.888)**	**0.830 (0.765)**

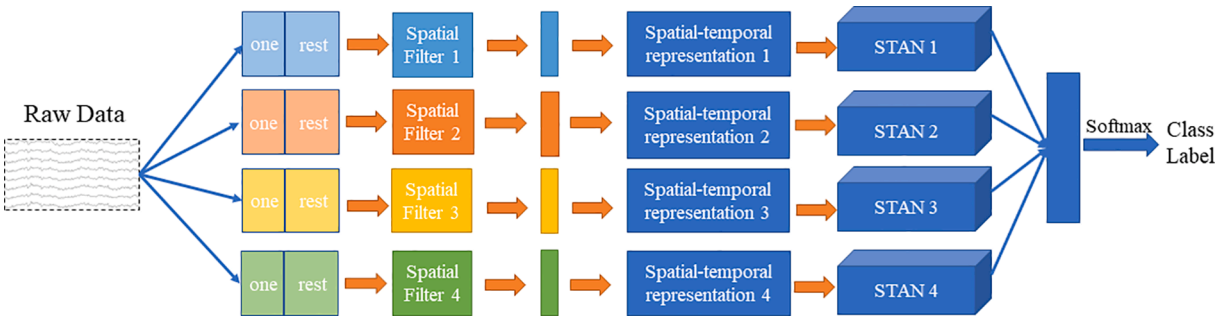

Fig. 6. Structure diagram of traditional OVR applied to our method.

Table 5
The results of the traditional OVR and Joint-OVR on the BCI IV 2a data set.

Accuracy (Kappa)	S1	S2	S3	S4	S5	S6	S7	S8	S9	
Traditional OVR	0.717 (0.635)	0.767 (0.682)	0.650 (0.612)	0.583 (0.463)	0.467 (0.397)	0.500 (0.328)	0.783 (0.531)	0.783 (0.682)	0.767 (0.627)	0.669 (0.551)
Joint-OVR	**0.883 (0.841)**	**0.817 (0.748)**	**0.922 (0.866)**	**0.776 (0.660)**	**0.633 (0.542)**	**0.675 (0.552)**	**0.900 (0.864)**	**0.950 (0.924)**	**0.917 (0.888)**	**0.830 (0.765)**

3.4.5. *Joint-OVR and traditional OVR strategy*

The Joint-OVR we proposed is different from the traditional OVR [28]. In the traditional OVR, multi-classification is also divided into multiple binary classifications. However, it does not merge these binary classifications, and each binary classification runs independently. Finally, it compares all of the results produced by each binary classification to predict the label. In other words, there are classifiers equal to the number of labels, and the final classification result is obtained by comparing the out probabilities of these classifiers. When the traditional OVR strategy is applied to our method instead of the Joint-OVR, the structure of our method becomes as shown in Fig. 6.

We compare the result of the traditional OVR and the Joint-OVR strategy applied to our method on the BCI IV 2a data set, and the result is shown in Table 5. From Table 5, we can find that the Joint-OVR has a better performance than the traditional OVR based on the method. The Joint-OVR is 0.161 higher ($p-value < 0.01$) in average accuracy than the traditional OVR. Separating each binary classification operation is unable to perform joint optimization, so that is not beneficial to classification. Moreover, we draw the statistical data in the form of the ROC line and AUC, as shown in Fig. 7. The curves of each category of the Joint-OVR are closer to the upper left corner, and the area under the curve is larger, which also reflects the better performance of the Joint-OVR. Besides, the traditional OVR needs 4 classifiers while the Joint-OVR only requires one. The network structure of the traditional OVR is four times larger than the Joint-OVR. It occupies a lot of server memory during training and consumes more training time. In summary, the Joint-OVR is more excellent than the traditional OVR.

3.4.6. *Compared with recent outstanding algorithms*

To verify that our method has a better performance than some different methods, we compare it with some outstanding algorithms on both BCI IV 2a and 2b data sets.

In Table 6, we evaluate the proposed method with Deep ConvNet [22] ($p-value < 0.01$), MCNN [23] ($p-value < 0.05$), C2CM [24] ($p-value = 0.01$), DCNN [33] ($p-value < 0.05$) and MS-AMF [34] ($p-value < 0.05$) on BCI IV 2a data set. We can see that our method obtains the highest accuracy rate in S2, S3, S6, S8, and S9, and the accuracy in other subjects is very close to the best result of others. The average accuracy rate of all subjects is the highest compared with other methods. In general, our method has more stable performance and higher classification accuracy on the BCI IV 2a data set.

We also evaluate the proposed method with the method proposed by

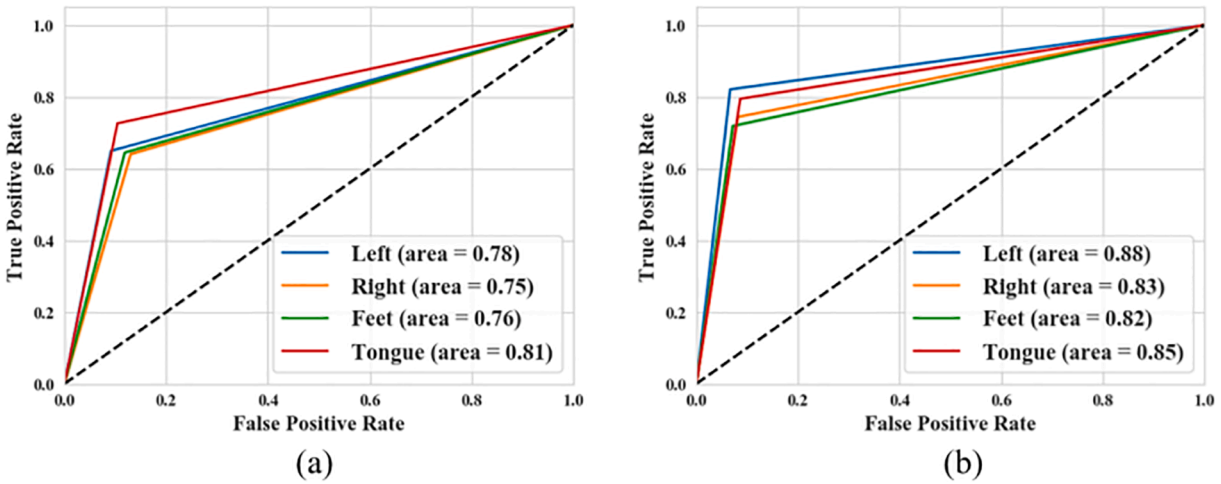

Fig. 7. ROC curve and AUC of traditional OVR and Joint-OVR classification results on the BCI IV 2a data set. (a) Traditional OVR; (b) Joint-OVR.

Table 6
The accuracy of some outstanding algorithms and our method on the BCI IV 2a data set.

Accuracy	S1	S2	S3	S4	S5	S6	S7	S8	S9	Average
Deep ConvNet [22]	0.866	0.623	0.899	0.656	0.552	0.485	0.861	0.784	0.761	0.721
MCNN [23]	**0.902**	0.634	0.894	0.712	0.628	0.477	**0.901**	0.837	0.823	0.757
C2CM [24]	0.875	0.653	0.903	0.667	0.625	0.455	0.896	0.833	0.795	0.745
DCNN [33]	0.803	0.7576	0.7442	0.5833	**0.7683**	0.6308	0.7841	0.8333	0.7821	0.743
MS-AMF [34]	0.8832	0.6559	0.9197	**0.7768**	0.6089	0.6367	0.8815	0.9323	0.8951	0.799
Our method	0.883	**0.817**	**0.922**	0.776	0.633	**0.675**	0.900	**0.950**	**0.917**	**0.830**

Table 7
The accuracy of some outstanding algorithms and our method on the BCI IV 2b data set.

Accuracy	S1	S2	S3	S4	S5	S6	S7	S8	S9	Average
Tang *et al.* [35]	0.8056	0.6544	0.6597	**0.9932**	0.8919	0.8611	0.8125	0.8882	0.8681	0.8261
BO [36]	0.714	0.6057	0.5809	0.9713	0.9132	0.8594	0.7698	0.9196	0.8453	0.7977
NCFS [37]	0.7925	0.6348	0.5665	0.9928	0.8867	0.7996	0.8876	0.9266	0.8495	0.8152
FDBN [38]	0.81	0.65	0.66	0.98	**0.93**	0.88	0.82	**0.94**	0.91	0.84
Our method	**0.861**	**0.779**	**0.676**	0.985	0.917	**0.958**	**0.917**	0.908	**0.917**	**0.880**

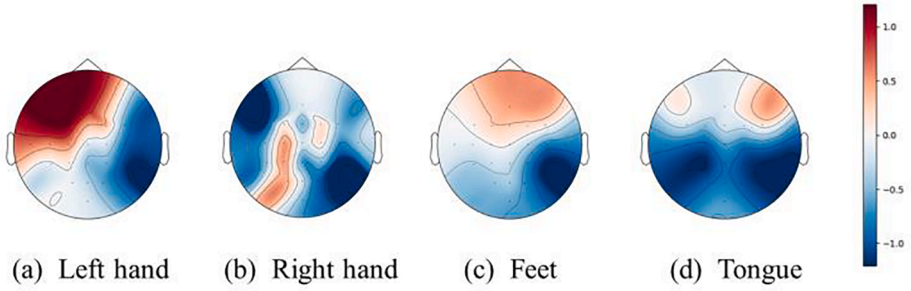

Fig. 8. The topographic map of Subject 8 reflects the different activation of the cerebral cortex by the subjects' different motor imagery.

Tang et al. [35] ($p-value < 0.05$), BO [36] ($p-value < 0.01$), NCFS [37] ($p-value < 0.05$) and FDBN [38] ($p-value < 0.05$) on the BCI IV 2b data set. From Table 7 our method achieves the highest accuracy rate in S1, S2, S3, S6, S7, and S9. The average classification accuracy rate is also the highest compared to other outstanding methods, reaching 0.880. Our method can still achieve a good classification effect even if applied to data with a small number of channels and binary classification tasks, showing generalization ability. These results prove that our method has a better performance compared with recent outstanding algorithms.

4. Discussion

Our proposed method is verified to extract discriminative spatial and temporal features efficiently, and has a good classification performance by several experiments. The TSF increases the inter-category difference of EEG and transforms the raw data into a spatial–temporal representation, which can be collaboratively optimized for multi-classification with the help of Joint-OVR. Subsequently, the STAN further extracts spatial and temporal representations from the spatial–temporal representation and classifies EEG.

The connectivity pattern of EEG indicates that its inter-channel relationships are functional and effective connectivity [39]. The relationship between an EEG channel and its adjacent channels is different from that between it and away from the channel. In the past, most methods directly stacked the EEG channels according to their serial numbers, which caused the receptive field during the convolution process to focus on the channels far away from each other spatially. Therefore, we consider the relationship between the channels and use the TSF based on covariance, a measure suitable for figuring out the relationship between every two channels. In Experiment B, the accuracy of TSF + STAN is 0.17 higher ($p-value < 0.01$) than it without the TSF, which also proves that this method is better than direct convolution.

On the other side, the different conditions of MI have relatively large differences under the influence of noise and field potential interference. We use the topographic map to show the different situations of the subject 8 of BCI IV 2a when performing motor imagery, as shown in Fig. 8, which shows that there is no apparent regular pattern. Even if the EEG we used is an ERS/ERD pattern, which indicates that there are differences in the activation of the left and right cerebral cortex, it is difficult to find discriminative features for distinguishing. The TSF uses the Joint-OVR strategy to map the EEG raw data to a spatial–temporal representation while increasing the inter-category difference. The extracted features from the spatial–temporal representation by the STAN are visualized by t-SNE [40] and shown in Fig. 9. We can see that TSF is beneficial to extract distinguishing features to improve classification performance.

In the process of spatial filtering, the TSF retains the time information, which is different from the original application of CSP. In Experiment C, compared with CSP, the TSF has a 0.099 ($p-value < 0.01$) higher accuracy rate, according to Table 3. Based on the same CNN, the CSP loses the temporal information of the raw data in the process of constructing feature vectors. On the contrary, the TSF maintains the time-series information, which allows the STAN to extract temporal features further to boost decoding.

The STAN utilizes the temporal characteristics and further strengthens the spatial characteristic based on the TSF. In the comparisons of Experiment D, the STAN outperforms other models in terms of feature extraction and classification performance. Both the spatial convolutional layer and the temporal convolutional layers in the ablation experiment indeed lead to a decrease in accuracy, which also shows that the STAN plays a role in the extraction of spatial and temporal features.

The Joint-OVR well extends our method from binary classification to

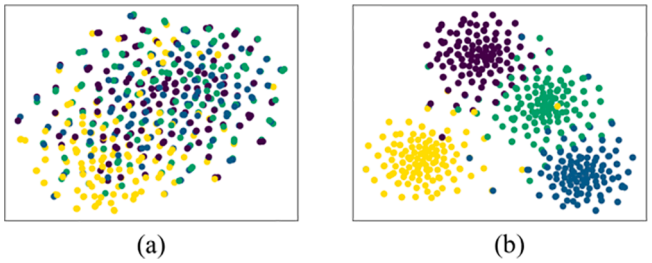

Fig. 9. Use t-SNE to visualize the comparison of the extracted features after the TSF processing and without the TSF processing. (a) The raw data; (b) The filtered data by the TSF.

multi-classification tasks. Besides, it inputs all the spatial–temporal representations after OVR into the STAN as a whole and makes them share the network, so that it can collaboratively optimize feature extraction in different spaces. The traditional OVR separates different spatial–temporal representations into different networks, and extracts features in their respective spaces. According to Table 5, the Joint-OVR has better classification performance than the traditional OVR.

However, there are shortcomings and limitations to our method currently. There is no detailed analysis of the impact of different parameters on the final results, despite the fact that we have done some pre-experiments to ensure the validity of the model. Besides, although we have worked hard to design the easy-STAN applicable to the CSP and TSF for comparison, it may not be the optimal model. Another is that the experiments we conducted are subject-specific. In the future work, we try to use this method to process cross-subject tasks, and explore more potential from it. It would be also valuable in feature research to see how it handles even spectral filtering. For instance, create spatial filters for each OVR in multiple frequency bands, in a similar way to the FBCSP [14], merge them all, filter the raw data, and input it into the neural network for collaborative optimization.

5. Conclusion

In this paper, we propose a feature extraction and classification method consisting of the TSF and the STAN, which can jointly utilize spatial and temporal features efficiently. Multiple mapping spaces are generated based on the number of EEG labels with TSF and then merged into one mapping space with the Joint-OVR. Then the STAN further explores discriminative spatial and temporal features, and finally achieves the purpose of decoding and classification. We design some experiments on the BCI competition IV-2a 4-class MI data set to verify that the TSF retains timing information while increasing inter-category difference, the STAN can extract discriminative spatial and temporal features, and the Joint-OVR strategy optimizes our approach. Besides, we evaluate our method with recent outstanding algorithms on the BCI IV 2a and 2b data sets, and get the highest classification accuracy.

CRediT authorship contribution statement

Xueyu Jia: Conceptualization, Methodology, Software, Validation, Data curation, Investigation, Resources, Writing – original draft, Visualization, Writing – review & editing. **Yonghao Song:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing. **Lie Yang:** Data curation, Investigation, Visualization. **Longhan Xie:** Supervision, Resources, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 52075177), Joint Fund of the Ministry of Education for Equipment Pre-Research (Grant No. 6141A02033124), Research Foundation of Guangdong Province (Grant No. 2019A050505001 and 2018KZDXM002), Guangzhou Research Foundation (Grant No. 202002030324 and 201903010028), Zhongshan Research Foundation (Grant No.2020B2020), and Shenzhen Institute of Artificial Intelligence and Robotics for Society (Grant No. AC01202005011).

References

[1] D. Zapała, E. Zabielska-Mendyk, P. Augustynowicz, A. Cudo, M. Jaśkiewicz, M. Szewczyk, N. Kopiś, P. Francuz, The effects of handedness on sensorimotor rhythm desynchronization and motor-imagery BCI control, Sci. Rep. 10 (1) (2020) 1–11.
[2] Y. Liu, W. Su, Z. Li, et al., Motor-imagery-based teleoperation of a dual-arm robot performing manipulation tasks, IEEE Trans. Cogn. Dev. Syst. 11 (3) (2018) 414–424.
[3] W.H. Lee, E. Kim, H.G. Seo, B.-M. Oh, H.S. Nam, Y.J. Kim, H.H. Lee, M.-G. Kang, S. Kim, M.S. Bang, Target-oriented motor imagery for grasping action: different characteristics of brain activation between kinesthetic and visual imagery, Sci. Rep. 9 (1) (2019) 1–14.
[4] M.A. Lebedev, M.A.L. Nicolelis, Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation, Physiol. Rev. 97 (2) (2017) 767–837.
[5] R. Mane, T. Chouhan, C. Guan, BCI for stroke rehabilitation: motor and beyond, J. Neural Eng. 17 (4) (2020) 041001, https://doi.org/10.1088/1741-2552/aba162.
[6] N. Cheng, K.S. Phua, H.S. Lai, P.K. Tam, K.Y. Tang, K.K. Cheng, R.-H. Yeow, K. K. Ang, C. Guan, J.H. Lim, Brain-computer interface-based soft robotic glove rehabilitation for stroke, IEEE Trans. Biomed. Eng. 67 (12) (2020) 3339–3351.
[7] T. Igasaki, J. Takemoto, K. Sakamoto, Relationship between kinesthetic/visual motor imagery difficulty and event-related desynchronization/synchronization, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 1911–1914.
[8] A.M. Savić, E.R. Lontis, N. Mrachacz-Kersting, M.B. Popović, Dynamics of movement-related cortical potentials and sensorimotor oscillations during palmar grasp movements, Eur. J. Neurosci. 51 (9) (2020) 1962–1970.
[9] B.A. Wang, S. Viswanathan, R.O. Abdollahi, N. Rosjat, S. Popovych, S. Daun, C. Grefkes, G.R. Fink, Frequency-specific modulation of connectivity in the ipsilateral sensorimotor cortex by different forms of movement initiation, NeuroImage 159 (2017) 248–260.
[10] M. Balbi, D. Xiao, M. Jativa Vega, H. Hu, M.P. Vanni, L.-P. Bernier, J. LeDue, B. MacVicar, T.H. Murphy, Gamma frequency activation of inhibitory neurons in the acute phase after stroke attenuates vascular and behavioral dysfunction, Cell Rep. 34 (5) (2021) 108696, https://doi.org/10.1016/j.celrep.2021.108696.
[11] M.T. Sadiq, X. Yu, Z. Yuan, F. Zeming, A.U. Rehman, I. Ullah, G. Li, G. Xiao, Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces, IEEE Access 7 (2019) 171431–171451.
[12] P.D. Purnamasari, T.W. Junika, Frequency-based EEG human concentration detection system methods with SVM classification, in: 2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom). IEEE, 2019: 29-34.
[13] G. Feng, L. Hao, G. Nuo, Feature extraction algorithm based on csp and wavelet packet for motor imagery EEG signals, in: 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), IEEE, 2019, pp. 798–802.
[14] K. Keng Ang, Z. Yang Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.), Hong Kong, Jun. 2008, pp. 2390–2397.
[15] Y. Guo, Y. Zhang, Z. Chen, Y.i. Liu, W. Chen, EEG classification by filter band component regularized common spatial pattern for motor imagery, Biomed. Signal Process. Control 59 (2020) 101917.
[16] R. Atangana, D. Tchiotsop, G. Kenne, et al., EEG signal classification using LDA and MLP classifier, Health Informat. Int. J. 9 (1) (2020) 14–32.
[17] Y. Narayan, E.E.G. Motor-Imagery, S.V.M. Signals Classificationusing, MLP and LDA Classifiers, Turkish J. Comput. Math. Educ. (TURCOMAT) 12 (2) (2021) 3339–3344.
[18] R. Chatterjee, T. Bandyopadhyay, D.K. Sanyal, et al., Comparative analysis of feature extraction techniques in motor imagery EEG signal classification, in: Proceedings of First International Conference on Smart System, Innovations and Computing, Springer, Singapore, 2018, pp. 73–83.
[19] Y. Narayan, Motor-imagery based EEG signals classification using MLP and KNN Classifiers, Turk. J. Comput. Math. Educ. (TURCOMAT) 12 (2) (2021) 3345–3350.
[20] A. Suleiman, Y.H. Chen, J. Emer, et al., Towards closing the energy gap between HOG and CNN features for embedded vision, in: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2017, pp. 1–4.
[21] T. Hori, S. Watanabe, Y. Zhang, et al. Advances in joint CTC-attention based end-to-end speech recognition with a deep CNN encoder and RNN-LM. arXiv preprint arXiv:1706.02737, 2017.

[22] T. Robin, Schirrmeister, et al., Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. (2017).
[23] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Mekhtiche, M. Shamim Hossain, Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion, Fut. Gener. Comput. Syst. 101 (2019) 542–554.
[24] S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks, IEEE Trans. Neural Networks Learn. Syst. 29 (11) (2018) 5619–5629.
[25] J. Kim, Y. Park, W. Chung, Transform based feature construction utilizing magnitude and phase for convolutional neural network in EEG signal classification, in: 2020 8th International Winter Conference on Brain-Computer Interface (BCI), IEEE, 2020, pp. 1–4.
[26] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, G. Pfurtscheller. BCI Competition 2008_Graz Data Set A. [Online]. Available: http://www.bbci.de/competition/iv/.
[27] R. Leeb, C. Brunner, G. Müller-Putz, et al., BCI Competition 2008–Graz data set B, Graz University of Technology, Austria, 2008, pp. 1–6.
[28] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Front. Neurosci. 6 (2012) 39.
[29] S.M. Vieira, U. Kaymak, J. Sousa, Cohen's kappa coefficient as a performance measure for feature selection, in: FUZZ-IEEE 2010, IEEE International Conference on Fuzzy Systems, Barcelona, Spain, 18-23 July, 2010, Proceedings. IEEE, 2010.
[30] P.L. Hsu, Contribution to the theory of" Student's" t-test as applied to the problem of two samples. Statistical Research Memoirs, 1938.
[31] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[32] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[33] E. Huang, X. Zheng, Y. Fang, Z. Zhang, Classification of motor imagery EEG based on time-domain and frequency-domain dual-stream convolutional neural network, IRBM (2021).
[34] D. Li, J. Xu, J. Wang, X. Fang, Y. Ji, A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (12) (2020) 2615–2626.
[35] X. Tang, W. Li, X. Li, W. Ma, X. Dang, Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network, Expert Syst. Appl. 149 (2020) 113285.
[36] H. Bashashati, R.K. Ward, A. Bashashati, User-customized brain computer interfaces using Bayesian optimization, J. Neural Eng. 13 (2) (2016) 026001.
[37] M.K.I. Molla, A.A. Shiam, M.R. Islam, T. Tanaka, Discriminative feature selection-based motor imagery classification using EEG signal, IEEE Access 8 (2020) 98255–98265.
[38] Q. Zheng, F. Zhu, P.-A. Heng, Robust support matrix machine for single trial EEG classification, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (3) (2018) 551–562.
[39] A.B. Buriro, R. Shoorangiz, S.J. Weddell, R.D. Jones, Predicting microsleep states using EEG inter-channel relationships, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (12) (2018) 2260–2269.
[40] P.E. Rauber, A.X. Falcão, A.C. Telea, Visualizing time-dependent data using dynamic t-SNE. 2016.