



## OPEN ACCESS

EDITED BY  
Jianjun Meng,  
Shanghai Jiao Tong University, China

REVIEWED BY  
Jing Jin,  
East China University of Science  
and Technology, China  
Kun Wang,  
Tianjin University, China

\*CORRESPONDENCE  
Yu Liu  
✉ liuyu54128@163.com

SPECIALTY SECTION  
This article was submitted to  
Brain-Computer Interfaces,  
a section of the journal  
Frontiers in Human Neuroscience

RECEIVED 13 October 2022  
ACCEPTED 05 December 2022  
PUBLISHED 21 December 2022

CITATION  
Feng J, Li Y, Jiang C, Liu Y, Li M and  
Hu Q (2022) Classification of motor  
imagery electroencephalogram  
signals by using adaptive  
cross-subject transfer learning.  
*Front. Hum. Neurosci.* 16:1068165.  
doi: 10.3389/fnhum.2022.1068165

COPYRIGHT  
© 2022 Feng, Li, Jiang, Liu, Li and Hu.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Classification of motor imagery electroencephalogram signals by using adaptive cross-subject transfer learning

Jin Feng<sup>2</sup>, Yunde Li<sup>1</sup>, Chengliang Jiang<sup>1</sup>, Yu Liu<sup>1\*</sup>, Mingxin Li<sup>3</sup>  
and Qinghui Hu<sup>4</sup>

<sup>1</sup>School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin, Guangxi, China, <sup>2</sup>Department of Student Affairs, Guilin Normal College, Guilin, Guangxi, China, <sup>3</sup>School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi, China, <sup>4</sup>School of Computer Science and Engineering, Guilin University of Aerospace Technology, Guilin, Guangxi, China

**Introduction:** Electroencephalogram (EEG)-based motor imagery (MI) classification is an important aspect in brain-computer interfaces (BCIs), which bridges between neural system and computer devices decoding brain signals into recognizable machine commands. However, due to the small number of training samples of MI electroencephalogram (MI-EEG) for a single subject and the great individual differences of MI-EEG among different subjects, the generalization and accuracy of the model on the specific MI task may be poor.

**Methods:** To solve these problems, an adaptive cross-subject transfer learning algorithm is proposed, which is based on kernel mean matching (KMM) and transfer learning adaptive boosting (TrAdaBoost) method. First, the common spatial pattern (CSP) is used to extract the spatial features. Then, in order to make the feature distribution more similar among different subjects, the KMM algorithm is used to compute a sample weight matrix for aligning the mean between source and target domains and reducing distribution differences among different subjects. Finally, the sample weight matrix from KMM is used as the initialization weight of TrAdaBoost, and then TrAdaBoost is used to adaptively select source domain samples that are closer to the target task distribution to assist in building a classification model.

**Results:** In order to verify the effectiveness and feasibility of the proposed method, the algorithm is applied to BCI Competition IV datasets and in-house datasets. The results show that the average classification accuracy of the proposed method on the public datasets is 89.1%, and the average classification accuracy on the in-house datasets is 80.4%.

**Discussion:** Compared with the existing methods, the proposed method effectively improves the classification accuracy of MI-EEG signals. At the same time, this paper also applies the proposed algorithm to the in-house dataset, the results verify the effectiveness of the algorithm again, and the results of this study have certain clinical guiding significance for brain rehabilitation.

## KEYWORDS

brain-computer interface, motor imagery, cross-subject transfer learning, kernel mean matching, TrAdaBoost

## 1 Introduction

Brain-computer interface (BCI) (Lin et al., 2021) is a technology, which directly establishes information interaction and control between the brain and external devices. BCI is able to translate neural responses into control instructions by decoding brain activity patterns from electroencephalogram (EEG) (Narayanan and Bertrand, 2019) signals. Motor Imagery (MI) (Tang et al., 2022) based BCI paradigm (MI-BCI) is one of the most popular paradigms nowadays. MI is defined as the cognitive process in which a person imagines their muscles or limbs moving without actually moving. Due to the above characteristics, the MI-BCI system has attracted wide attention in fields as stroke rehabilitation, wheelchair, or prosthetic control, etc. For example, stroke patients can achieve rehabilitation training through MI (Wang et al., 2019) to improve their motor function and restore the ability to control body parts. Therefore, accurate identification and classification prediction of MI-BCI signals is of great significance in the field of medical rehabilitation.

However, generally the number of subjects participating in the experiment is limited, which causes the dataset to be of small sized. In addition, there are great individual differences between subjects and between sessions (Saini et al., 2020), and it is difficult to directly use the MI-EEG signals across subjects to construct a robust classification model. Thus, how to effectively distinguish the types of MI-EEG signals and achieve high performances with small-sized datasets has become the key challenge of MI-BCI system. Transfer learning improves model performance in target tasks by learning and transferring information from source tasks. Some researchers hope to reduce the individual differences in the distribution of MI-EEG signals through transfer learning (Cao et al., 2021; Xu et al., 2021), and build a classification model with high robustness on the MI-EEG signals of all subjects.

Transfer learning mainly includes three learning paradigms: parameter transfer learning, feature transfer learning, and instance transfer learning. For the parameter transfer learning, there is a supervised weighted logistic regression-based transfer learning (S-wLTL), which added the regularization parameters to the objective function of the classifier to make the classification parameters as close as possible to other similar subjects. The classification accuracy of this method on BCIIV dataset IIA reaches 75.6%, which is higher than that of common spatial pattern (CSP) combined with support vector machine (SVM) (Azab et al., 2019). Based on the MI-EEG signals of other source subjects, a parallel multiscale filter bank convolutional neural network is pre-trained, and then fine-tuned in the individual target task. The experimental results show that the classification accuracy of cross-subject classification reaches 75.9% (Wu et al., 2019). However, when the parameters of the source domain are transferred to the target domain, the catastrophic forgetting problem may occur with the iterative

optimization of the algorithm, which leads to low accuracy of the results obtained by these methods (Guo et al., 2021).

For the feature transfer learning, the algorithm based on the maximum mean difference (MMD) (Pan et al., 2010) is the common method. MMD is used to select features that are more relevant to the target task by measuring the distribution difference between source and target domain samples. Li et al. (2020) proposed a discriminative transfer feature learning (DTFL), which enhances class discrimination information by minimizing the marginal and condition distribution between the source and target domain, while maximizing the distance between the classes. The experimental results show that classification accuracy is 83.5%. In addition, a weighted logistic regression method (Chen et al., 2022) based on Euclidean feature space is proposed, and the MI-EEG signals of different subjects are aligned in Euclidean space (He and Wu, 2019) to reduce the difference of signals. Then the Kullback–Leibler (KL) divergence (Shibanoki et al., 2018) of CSP features between different subjects is calculated in Euclidean space, and the average classification accuracy of this method is 85%. The feature transfer learning method reduces the distribution differences between the source and target domain by specific data distribution analysis methods. However, when the data distribution between source and target domain samples are very different, the effect of feature transfer learning will be seriously affected, and even may cause negative transfer (Wan et al., 2021).

For the instance transfer learning, the main idea is to train the target classifier by selecting the source domain samples with more similar distribution to the target samples. For examples, Tan et al. (2018) proposes an instance transfer ensemble learning for Alzheimer's disease classification, which is used to select and transfer source domain samples with similar to target domain samples by instance transfer learning algorithm (ITL) based on wrapper mode, thereby obtaining optimal transferred domain samples. Dai et al. (2007) proposed a boost transfer learning (TrAdaBoost) algorithm, which applied the idea of adaptive boosting (Adaboost) (Chen et al., 2021) in transfer learning to improve the instance weight of the target classification task and reduce the instance weight of the unfavorable target task. Similarly, Huang et al. (2007) proposed a kernel mean matching (KMM) algorithm, which aims to make the probability distribution of the weighted source domain sample and the target domain sample as close as possible.

Inspired by the above studies, this study combines KMM and TrAdaBoost algorithms to analyze MI-EEG signals for the first time and proposes a KMM-TrAdaBoost instance transfer learning algorithm. The KMM-TrAdaBoost is mainly used to solve the problem that the small number of training samples of MI-EEG for a single subject and the great individual differences of MI-EEG signals among different subject resulting the poor performance of the classification model. Firstly, the algorithm preprocesses the source and target domain samples to extract their CSP features. Then uses the KMM algorithm to obtain the sample weight matrix between source and target domain,

which is used as the initial weight matrix of TrAdaBoost to solve the problem that TrAdaBoost algorithm is sensitive to the initialization weight. Finally, the strong classifier, which integrates multiple weak classifiers based on TrAdaBoost is used for classification. The KMM-TrAdaBoost algorithm proposed in this paper can effectively improve the classification accuracy for a single subject in the case of small samples. At the same time, the model also showed good performance on the samples of different subjects, and obtained high classification accuracy and generalization performance, which reduced the influence of individual differences on MI-EEG classification. In addition, this method has certain clinical value for cross-individual rehabilitation therapy.

## 2 Methodology

### 2.1 Overview

The flow chart of proposed KMM-TrAdaBoost is shown in **Figure 1**. Firstly, the KMM algorithm is used to calculate the weights matrix of all training sample to estimate the similarity between the source and target sample. Then the matrix is used as the initialization weight matrix of TrAdaBoost algorithm. Finally, the strong classifier, which integrates multiple weak classifiers based on TrAdaBoost is used for classification. The specific algorithm steps are as follows:

- (1) **Preprocessing all source and target domain data.** The preprocessing steps comprise band-pass filtering, time window processing, and channel selection.
- (2) **Spatial features extracting.** Spatial features are extracted from the preprocessed data by using CSP algorithm.
- (3) **Calculating sample weight matrix.** The KMM algorithm is used to obtain a sample weight matrix, which makes the mean values of the features between the source and target domain as close as possible.
- (4) **Training several weak classifiers.** The sample weight matrix, which is calculated by KMM algorithm, is as an initialization sample weight matrix of the TrAdaBoost algorithm. Then, several weak classifiers are trained based on the weighted samples.
- (5) **Obtaining strong classifier by voting strategy.** According to the weight of each classifier, the final strong classifier is obtained by using the voting strategy.

### 2.2 CSP algorithm

Common spatial pattern (Ang et al., 2012) is a spatial filtering feature extraction algorithm for two-class MI-EEG tasks, which aims to use matrix diagonalization to find a set of optimal spatial filters and project them to maximize the variance difference between the two class of signals. And

then, the specificity feature vectors with high discrimination were obtained to achieve the task of distinguishing two kinds of MI-EEG signal. The CSP algorithm flow is shown in **Figure 2**.

Suppose  $X_1$  and  $X_2$  are the time-space signal matrices of multichannel evoked responses under the dichotomous MI task, and their dimensions are  $N \times T$ .  $N$  are the number of channels of the EEG device, and  $T$  is the number of sampling points of each channel.  $X_1$  and  $X_2$  can be expressed as:

$$X_1 = [C_1 \ C_M] \begin{bmatrix} S_1 \\ S_M \end{bmatrix} \quad X_2 = [C_2 \ C_M] \begin{bmatrix} S_2 \\ S_M \end{bmatrix} \quad (1)$$

Where  $S_1$  and  $S_2$  represents two classification MI tasks in source domain, and they are assumed to be linearly independent of each other.  $S_M$  represents a source signal that is common to both types of tasks, assuming that  $S_1$  consists of  $m_1$  source domain signals and  $S_2$  consists of  $m_2$  source domain signals. Then,  $C_1$  and  $C_2$  are composed of  $m_1$  and  $m_2$  common spatial patterns associated with  $S_1$  and  $S_2$ , respectively. Each spatial pattern is a  $N \times 1$  dimensional vector, which represents the distribution weight of the signal caused by a single source signal on  $N$  leads.  $C_M$  represents the corresponding shared spatial pattern with  $S_M$ . The goal of CSP algorithm is to design a spatial filter to obtain the spatial factor  $W$ , and its process is shown as follows. (1) Normalizing  $X_1$  and  $X_2$ , respectively, their corresponding covariance matrices  $R_1$  and  $R_2$  are calculate. They can be expressed as:

$$R_1 = \frac{X_1 X_1^T}{\text{trace}(X_1 X_1^T)} \quad R_2 = \frac{X_2 X_2^T}{\text{trace}(X_2 X_2^T)} \quad (2)$$

Where  $\text{trace}(\bullet)$  is the sum of the entries on the diagonal for the matrix, and  $T$  is the transpose of the matrix. (2) The mixed spatial covariance  $R$  from  $R_1$  and  $R_2$  can be expressed as:

$$R = \overline{R_1} + \overline{R_2} \quad (3)$$

Where  $\overline{R_1}$  and  $\overline{R_2}$  are the average covariance matrices, respectively.

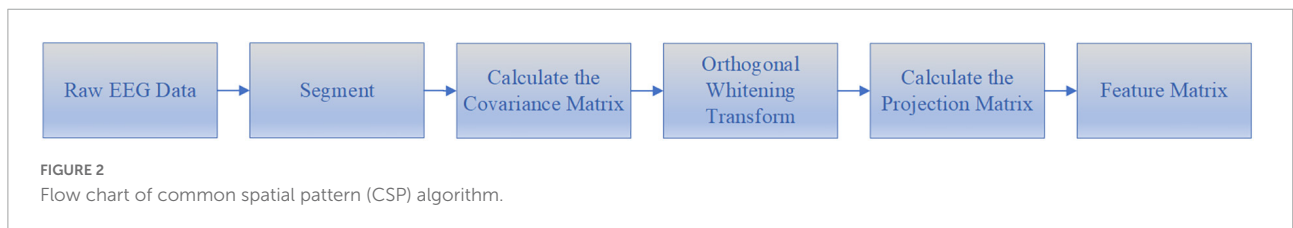
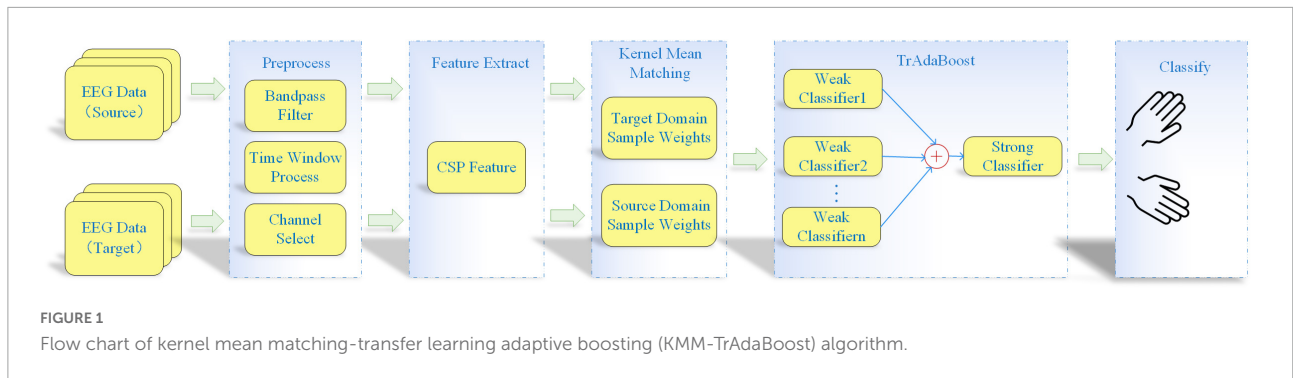
Based on the eigenvalue decomposition theory,  $R$  is expressed as  $R = U \lambda U^T$ , where  $U$  is an eigenvector of the matrix,  $\lambda$  is the corresponding eigenvalue. In addition, the eigenvalues are arranged in descending order, and the corresponding eigenvectors are also rearranged. Then, principal component analysis theory (Ahuja et al., 2022) was used to calculate the whitening matrix  $P$ , and it be expressed as:

$$P = \sqrt{\lambda^{-1}} U^T \quad (4)$$

(3) Based on the whitening matrix  $P$ , the covariance matrix  $R_1$  and  $R_2$  can be transformation, and it is expressed as:

$$S_1 = P R_1 P^T \quad S_2 = P R_2 P^T \quad (5)$$

Where  $S_1$  and  $S_2$  have a common eigenvector. if  $S_1 = B \lambda_1 B^T$ , then  $S_2 = B \lambda_2 B^T$  and  $\lambda_1 + \lambda_2 = I$ , where  $B$  is common eigenvector of  $S_1$  and  $S_2$ , and  $I$  is the identity matrix. Since



the eigenvalues of the two matrices always add up to 1, the eigenvector of  $S_1$  corresponding to the largest eigenvalue cause  $S_2$  to have the smallest eigenvalue.

The projection matrix  $W$  is the corresponding spatial filter, and it be expressed as  $W = B^T P$ . Then the EEG data of the single task experiment  $X_i$  can be transformed into  $Z = W^* X_i$   $i = (1, 2)$ . For the EEG signals, the feature value  $f_p$  can be expressed as:

$$f_p = \lg \left[ \frac{\text{var}(Z_p)}{\text{sum}(\text{var}(Z_p))} \right] \quad p = (1, 2, 3, \dots, 2m) \quad (6)$$

### 2.3 KMM-TrAdaBoost algorithm

#### 2.3.1 KMM algorithm

Instance transfer learning is to select samples from the source domain, which are consistent with the distribution of the target samples, and improves model performance in target tasks by transferring information from source domain. The KMM (Huang et al., 2007) algorithm maps the source domain samples from original feature space into reproducing kernel Hilbert space (RKHS) (Gertton et al., 2012), and then calculates the difference between the mean value of the source and target domain data under the RKHS space. Finally, a set of weight parameters matrix are obtained, which are used to weight the samples in the source domain to make the probability distribution consistent with the samples in the target domain. The calculation process of KMM algorithm is as follows.

$$\min \left\| \frac{1}{m} \sum_{i=1}^m \beta_i \phi(x_i^s) - \frac{1}{n} \sum_{i=1}^n \phi(x_i^t) \right\|_H^2 \quad (7)$$

Where  $x_i^s$  is a set of source domain sample ( $i = 1, 2, \dots, m$ ), and  $x_i^t$  is a set of target domain sample ( $i = 1, 2, \dots, n$ ).  $H$  denotes the RKHS with a characteristic kernel  $k$ .  $\beta_i \in [0, 1]$  represents the weight of the  $i$ -th source domain sample.  $\phi(\bullet)$  is the mapping function from the original space to the RKHS, and satisfies the following relation:  $\langle \phi(x), \phi(y) \rangle_H = k(x, y)$ .  $k(x, y)$  is a Gaussian kernel function, namely:

$$k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (8)$$

where  $\sigma$  represents the size of the Gaussian kernel. Combining Equations 7, 8, the MMD between each source and target domain is defined as:

$$\min \left( \frac{1}{m^2} \left( \sum_{i=1}^m \beta_i \phi(x_i^s) \right)^2 - \frac{2}{nm} \sum_{i=1}^m \beta_i \phi(x_i^s) \sum_{i=1}^n \phi(x_i^t) + c \right) \quad (9)$$

where  $c$  stands for constant, and  $\sum_{i=1}^m \beta_i \phi(x_i^s)^2$  and  $\sum_{i=1}^n \beta_i \phi(x_i^s) \sum_{i=1}^n \phi(x_i^t)$  can be reduced to a matrix form as:

$$\begin{aligned} & \left( \sum_{i=1}^m \beta_i \phi(x_i^s) \right)^2 \\ &= (\beta_1 \phi(x_1^s) + \dots + \beta_m \phi(x_m^s))^2 \\ & \quad k(x_1^s, x_2^s) \dots k(x_1^s, x_m^s) \\ &= [\beta_1 \ \beta_2 \dots \beta_m]^* \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} [\beta_1 \ \beta_2 \dots \beta_m]^T = \beta^T K \end{aligned} \quad (10)$$

$$\sum_{i=1}^m \beta_i \phi(x_i^s) \sum_{i=1}^n \phi(x_i^t)$$

$$= [\sum_{j=1}^n k(x_1^s, x_j^t) \dots \sum_{j=1}^n k(x_n^s, x_j^t)] [\beta_1 \beta_2 \dots \beta_m]^T = g^T \beta \tag{11}$$

Where  $K_{i,j} = k(x_i^s, x_j^t)$  and  $g_j = \frac{m}{n} \sum_{i=1}^n k(x_i^s, x_j^t)$ . Then the final quadratic optimization objective function is as follows.

$$\min(\frac{1}{n^2} \beta^T K \beta - \frac{1}{m^2} g^T \beta) \tag{12}$$

### 2.3.2 TrAdaBoost algorithm

Inspired by the algorithm of AdaBoost (Chen et al., 2021), TrAdaBoost (Dai et al., 2007) uses weight automatic updating mechanism to constantly adjust the weight of samples, so as to keep important source domain samples and eliminate the samples that are not similar to the distribution of target domain samples. The principle of the TrAdaBoost algorithm is shown in Figure 3.

$D^s = \{(x_i^s, y_i^s)\}_{i=1}^n$  represents samples from source domain, where  $x_i^s \in D^s (i = 1, \dots, n)$  is a sample example,  $y_i^s$  is the corresponding ground-truth labels.  $D^t = \{(x_j^t, y_j^t)\}_{j=1}^m$

represents samples from target domain, where  $x_j^t \in D^t (j = 1, \dots, m)$  is a sample example,  $y_j^t$  is the corresponding ground-truth labels.  $n$  and  $m$  represent the sizes of  $D^s$  and  $D^t$ .  $T^t$  is the test samples from target domain, which is assumed to follow a different probability distribution from  $D^s$  and the same probability distribution as  $D^t$ . The goal of transfer learning is to train a classifier with the minimum classification error based on a small amount of target domain samples  $D^t$  and a large amount of source domain samples  $D^s$ . The specific algorithm flow is shown as follows.

(1) Initialize samples weight  $w^1 = (w_1^1, \dots, w_{n+m}^1)$  from source and target domain samples  $D = \{D^s, D^t\}$ :

$$w_k^1 = \frac{1}{n+m}, k = 1, \dots, n+m \tag{13}$$

(2) Set initial parameter  $\beta$  of source domain sample:

$$\beta = 1/(1 + \sqrt{2 \ln n/N}) \tag{14}$$

Where  $N$  is the number of iterations.

(3) Normalizing the weight vector  $w$ , and a weak classifier  $h_1$  is trained based on the weighted sample  $D$  and test data  $T^t$ , and

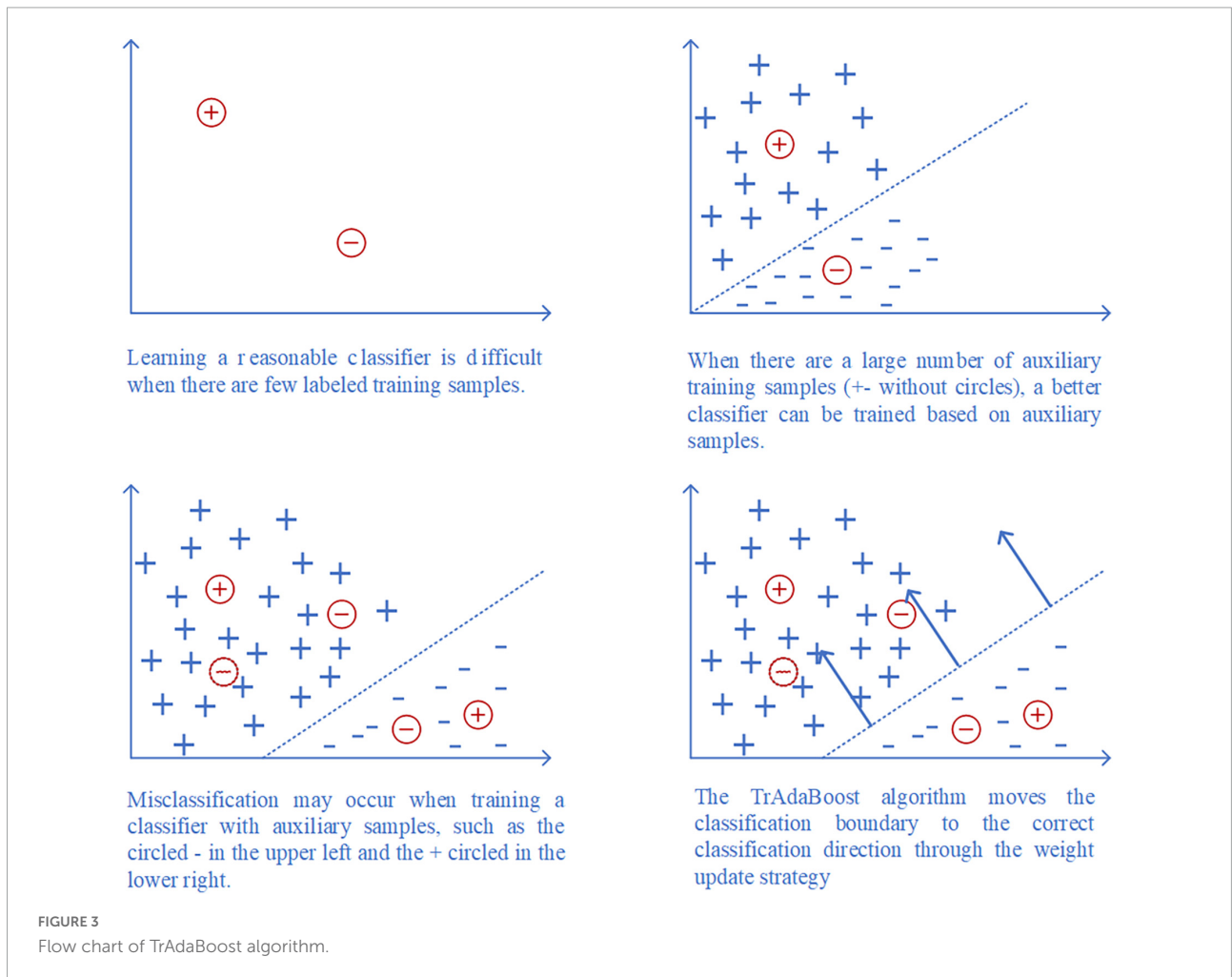


FIGURE 3  
Flow chart of TrAdaBoost algorithm.

SVM is adopted as a weak classifier in this paper. Calculate the error of the weak classifier  $h_l$  on the training dataset of the target domain, and the calculation formula is as follows.

$$\varepsilon_l = \sum_{i=n+1}^{n+m} \frac{w_i^l |h_l(x_i^t) - y_i^t|}{\sum_{i=n+1}^{n+m} w_i^l} \quad (15)$$

Where  $l = 1, \dots, N$  represents the  $l$ -th iteration.

(4) Set up  $\beta_l = \varepsilon_l / (1 - \varepsilon_l)$ , and update weight

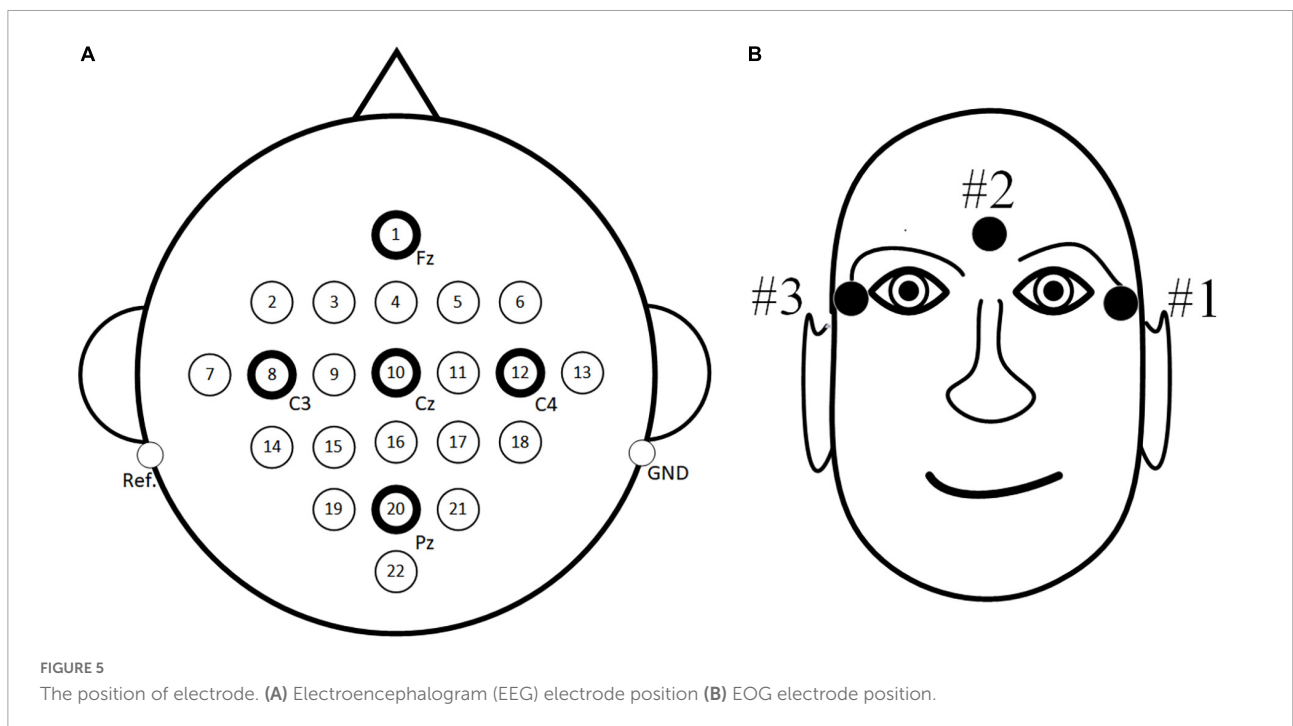
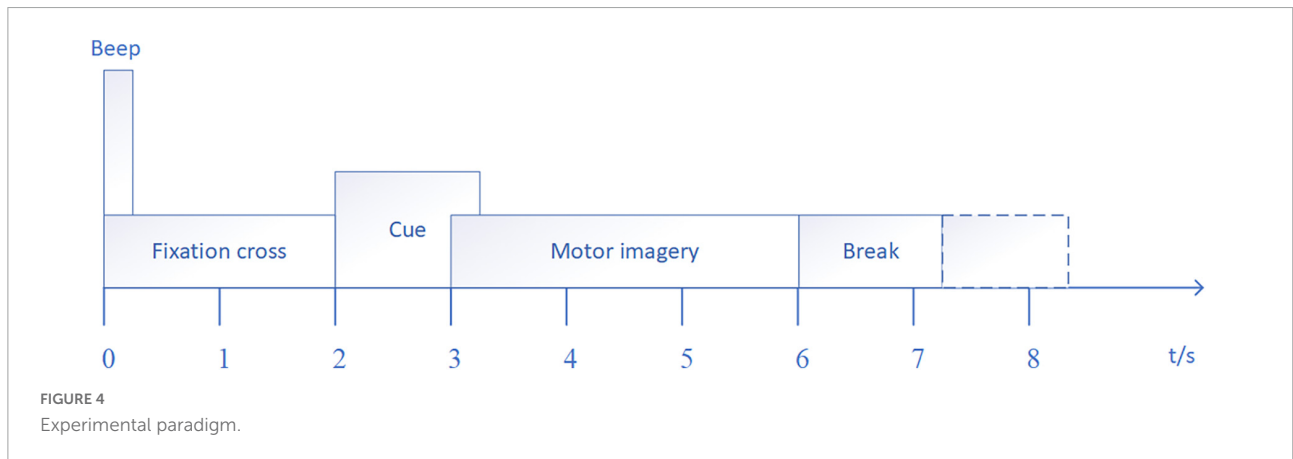
$$w_i^{l+1} = \begin{cases} w_i^l \beta_l^{|h_l(x_i) - y_i^t|} & i = 1, 2, 3, \dots, n \\ w_i^l \beta_l^{-|h_l(x_i) - y_i^t|} & i = n + 1, \dots, n + m \end{cases} \quad (16)$$

(5) Repeat steps (3–4) and iterate for  $N$  times to obtain the final strong classifier

$$h(x) = \begin{cases} 1, & \sum_{l=\lceil N/2 \rceil}^N \ln(1/\beta_l) h_l(x) \geq \frac{1}{2} \sum_{l=\lceil N/2 \rceil}^N \ln(1/\beta_l) \\ 0, & \text{other} \end{cases} \quad (17)$$

## 2.4 Statistical analysis for different datasets

For different datasets, the two-tailed  $t$ -test (Xie et al., 2022) was used to compare the significant difference between the proposed method and state-of-the-art methods, and  $P < 0.05$



was considered statistically significant. All statistical tests were carried out with Origin 2022 software.

## 3 Result and discussion

### 3.1 Dataset

#### 3.1.1 Public dataset

For the public dataset 1, the BCI Competition IV dataset 2a (Blankertz et al., 2006) is used for research. The dataset consisted of EEG signals recorded by 22 electrodes from nine subjects with three EOG scalp electrode locations. The dataset consists of four tasks, which are left hand, right hand, foot and tongue MI task. 72 MI experiments were performed for each task, and a total of 288 experiments. The signal sampling frequency was 250 Hz, and the band pass filter between 0.5–100 Hz and notch filter of 50 Hz are used to eliminate the power frequency interference. In order to be consistent with the in-house dataset, only the MI data of the left and right hands MI task are used in this paper. Specific experimental paradigms are shown in Figure 4, and EEG electrode positions are shown in Figure 5A as follows: Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, and POz, left mastoid reference, right mastoid grounding. The position of EOG electrode is shown in Figure 5B.

For the public dataset 2, BCI Competition IV dataset I (Blankertz et al., 2006) is adopted for research. The data set was EEG signals recorded by 59 electrodes in 7 subjects. There are three categories of MI tasks, namely left and right hand and foot imagination. The sampling frequency is 100 Hz. In order to ensure the consistency of experimental paradigm, left and right hand categories are extracted from three motor imagination tasks in this paper, and each category of each user has 100 samples. Electrode position adopts international 10/20 system.

#### 3.1.2 In-house dataset

An in-house dataset was recorded from 20 subjects (S01-S20) who are divided into two groups, including ten (S01-S10) who are good at table tennis and 10 (S11-S20) who are not. Scan software, Neuroscan37 Quik-Cap electrode cap and SynAmps2 amplifier, and MATLAB software were used for data acquisition and analysis. The prompting mode of “preparation-prompt-movement” was adopted. The prompting was divided into two ways: imaginary right-handed attack and left-handed attack. Each player did 144 experiments. The dataset collected in this paper consists of 32 electrodes, A1 and A2 are used as reference electrodes, and EEG signals are collected from the scalp of each subject at a sampling frequency of 250 Hz. “HEOL” and “HEOR” is used to obtain the horizontal electro-oculogram. “VEOU” and “VEOL” is used to obtain the vertical electro-oculogram.

TABLE 1 Results of comparison with state-of-the-art methods on the public dataset 1 and public dataset 2.

	Method	Acc			K		
		EA-CSP-LDA	CA-JDA	Ours	EA-CSP-LDA	CA-JDA	Ours
Public dataset 1	A01	86.2%	65.8%	<b>92.5%</b>	0.72	0.43	<b>0.81</b>
	A02	58.6%	51.8%	<b>89.7%</b>	0.43	0.35	<b>0.75</b>
	A03	96.6%	65.9%	<b>92.6%</b>	0.90	0.44	<b>0.83</b>
	A04	71.6%	62.3%	<b>81.9%</b>	0.57	0.46	<b>0.65</b>
	A05	54.6%	54.8%	<b>90.2%</b>	0.36	0.40	<b>0.79</b>
	A06	66.9%	58.6%	<b>85.6%</b>	0.46	0.42	<b>0.69</b>
	A07	68.9%	67.2%	<b>91.7%</b>	0.48	0.47	<b>0.78</b>
	A08	86.9%	88.2%	<b>89.9%</b>	0.73	0.70	<b>0.71</b>
	A09	77.9%	71.9%	<b>87.9%</b>	0.60	0.58	<b>0.76</b>
		Average ± Std	(74.2 ± 12.3)%	(65.1 ± 9.5)%	<b>(89.1 ± 3.1)%</b>	0.58 ± 0.16	0.47 ± 0.09
Public dataset 2	A01	75.8%	62.8%	<b>89.5%</b>	0.59	0.46	<b>0.75</b>
	A02	75.9%	59.9%	<b>83.5%</b>	0.59	0.39	<b>0.67</b>
	A03	74.8%	60.9%	<b>84.6%</b>	0.58	0.40	<b>0.69</b>
	A04	73.5%	61.3%	<b>85.1%</b>	0.56	0.41	<b>0.70</b>
	A05	90.6%	68.6%	<b>90.8%</b>	0.79	0.48	<b>0.81</b>
	A06	77.8%	67.8%	<b>80.9%</b>	0.61	0.47	<b>0.60</b>
	A07	85.9%	61.8%	<b>84.2%</b>	0.71	0.41	<b>0.62</b>
		Average ± Std	(79.1 ± 5.6)%	(63.3 ± 3.0)%	<b>(85.5 ± 2.9)%</b>	0.63 ± 0.07	0.43 ± 0.03

Std stands for standard deviation. Bold indicates the maximum value in the current indicator.

### 3.2 Data preprocessing

In the data preprocessing stage, 3 EOG channels were deleted and 22 EEG channels are retained for the public dataset. Different from the public dataset, the in-house dataset deletes the 6 electrode data of “HEOL,” “HEOR,” “A1,” “A2,” “VEOU,” and “VEOL.” 26 EEG channels are retained to reduce the false recording caused by the EOG signal. According to the principle of event related desynchronization (ERD) and event related synchronization (ERS) (Tang et al., 2022), when people perform MI tasks, the cerebral cortex will produce obvious rhythm signals, which are divided into 8–12 Hz  $\mu$  rhythm signals and 13–30 Hz  $\beta$  rhythm signals. In order to improve the signal-to-noise ratio of EEG signals, 8–30 Hz band-pass filtering is used to process the data and remove the baseline. At the same time, the corresponding EEG signal was extracted as the subsequent analysis data by using the time window of 0.5–3.5 s after the appearance of the prompt.

The MI-EEG data used in this paper consists of two parts, including nine subjects from the IV2a dataset (A01, A02, ..., A09)

TABLE 2 Comparison of the accuracies (%) between the proposed algorithm and other methods on the public dataset 1 and public dataset 2.

	Target	CSP+KMM	CSP+TrAdaboost	CSP+SVM	Ours
Public dataset 1	A01	83.3%	75%	60.7%	<b>92.5%</b>
	A02	87.3%	78.4%	59.6%	<b>89.7%</b>
	A03	86%	76.8%	63.5%	<b>92.6%</b>
	A04	78.4%	72.3%	53.9%	<b>81.9%</b>
	A05	85.6%	80.4%	54.6%	<b>90.2%</b>
	A06	78.3%	72.6%	67.2%	<b>85.6%</b>
	A07	87.8%	84.3%	54.3%	<b>91.7%</b>
	A08	85.6%	81.4%	58.2%	<b>89.9%</b>
	A09	84.2%	80.3%	54.9%	<b>87.9%</b>
	Average $\pm$ Std	(84.1 $\pm$ 3.1)%	(77.9 $\pm$ 3.6)%	(58.5 $\pm$ 4.1)%	<b>(89.1 <math>\pm</math> 3.1)%</b>
Public dataset 2	A01	73.3%	78.6%	60.4%	<b>89.5%</b>
	A02	77.3%	79.4%	57.6%	<b>83.5%</b>
	A03	76.5%	78.8%	53.5%	<b>84.6%</b>
	A04	74.4%	76.7%	58.9%	<b>85.1%</b>
	A05	72.6%	80.4%	54.9%	<b>90.8%</b>
	A06	74.3%	72.8%	58.2%	<b>80.9%</b>
	A07	72.8%	84.3%	54.3%	<b>84.2%</b>
		Average $\pm$ Std	(74.4 $\pm$ 1.5)%	(78.7 $\pm$ 3.0)%	(56.8 $\pm$ 2.2)%

Std stands for standard deviation. Bold indicates the maximum value in the current indicator.

and twenty subjects (S01-S20) from the in-house dataset. For the IV2a dataset, the EEG data of one subject was selected as the target domain data, and the data of the other eight subjects was selected as the source domain data. For the 20 subjects from the in-house dataset, they were divided into two groups: 10 subjects who could play ping-pong and 10 subjects who could not. There were 10 subjects in each group. One subject from the same group was selected as the target domain data, and the remaining 9 subjects were used as the source domain data. In each set of data, 70% of the data is the training set and 30% of the data is the test set.

### 3.3 Result evaluation index

To evaluate the performance of the classification model, the accuracy (Acc) and Kappa value (K) are measured, and the calculation equations are as follows:

$$\begin{cases} Acc = \frac{TP+TN}{TP+FN+TN+FP} \\ K = \frac{Acc-P_e}{1-P_e} \end{cases} \quad (18)$$

Where TP is the number of samples correctly classified as positive label, TN is the number of samples correctly classified as negative label, FP is the number of samples misclassified as positive label, and FN is the number of samples misclassified as negative label.  $P_e = (a_1 \times b_1 + a_2 \times b_2 + \dots + a_z \times b_z) / n^2$  represents the random classification rate of model to samples.  $a_1, a_2, \dots, a_z$  represents the actual sample size of each type of sample, and  $b_1, b_2, \dots, b_z$  represents the number of samples of each type predicted by the model  $n$  is the total number of samples. Through the comparative analysis of Kappa values, the influence of random classification on the accuracy of the model can be eliminated.

### 3.4 Analysis of experimental results

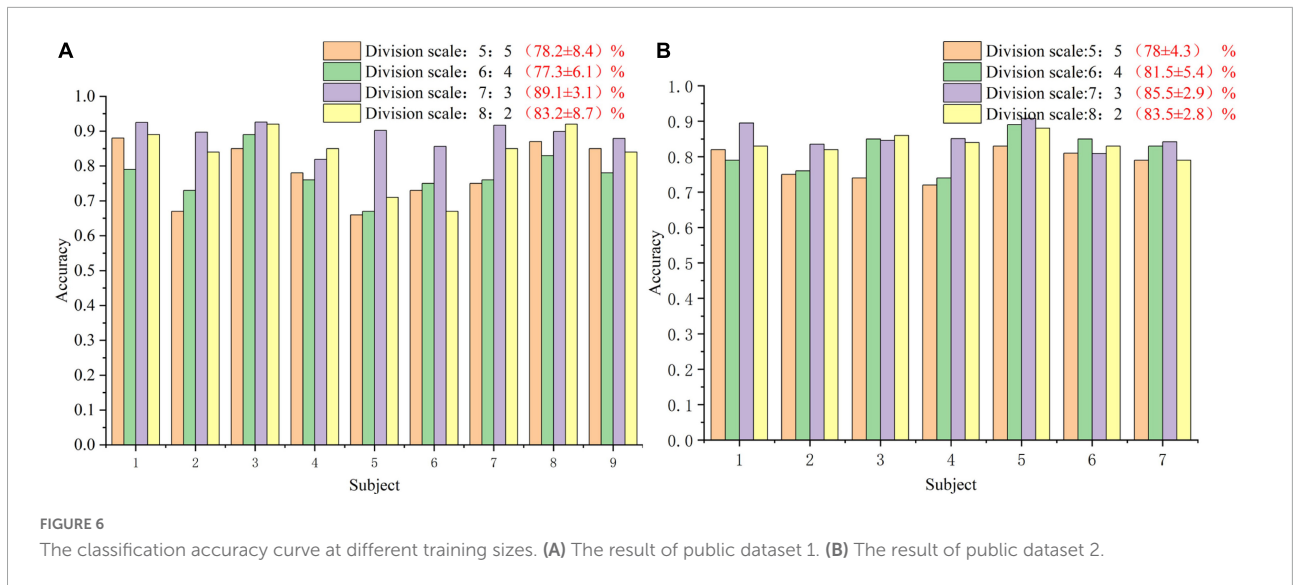
#### 3.4.1 Comparison with state-of-the-art methods

In order to verify the effectiveness of the proposed algorithm (Ours), the following two methods are compared, including the EA-CSP-LDA (He and Wu, 2019) and CA-JDA (Iturrate et al., 2013) algorithm.

The EA-CSP-LDA algorithm reduces the difference of cross-subject signals based on the Euclidean space data alignment approach. For the CA-JDA algorithm, joint distribution adaptation (JDA) is used to align the edge probability distribution and conditional probability distribution of cross-subject signals to achieve effective transfer learning. From these results in Table 1, we can obtain the following insightful observations.

(1) Compared with the proposed method, the EA-CSP-LDA and CA-JDA algorithms do not take into account the samples





with large differences, especially for the recognition of cross-individual MI-EEG signals, so the performance of the above algorithms is poor.

(2) The proposed method obtains the best prediction results on the MI-EEG classification tasks, verifying its effectiveness and superiority.

TABLE 3 T-test results for the proposed method vs. state-of-the-art methods.

Paired T-test	P-value	
	Public dataset 1	Public dataset 2
Ours vs. EA-CSP-LDA	*	*
Ours vs. CA-JDA	*	**

~ nonsignificant, \* $P \leq 0.05$ , \*\* $P \leq 0.01$ .

TABLE 4 Separate session of T-test statistical analysis for the proposed method vs. state-of-the-art methods.

Paired T-test	P-value	
	Ours vs. EA-CSP-LDA	Ours vs. CA-JDA
A01	*	**
A02	*	*
A03	~	**
A04	*	*
A05	*	*
A06	*	*
A07	*	*
A08	~	~
A09	*	*
Average	*	*

~ nonsignificant, \* $P \leq 0.05$ , \*\* $P \leq 0.01$ .

### 3.4.2 Ablation experiment

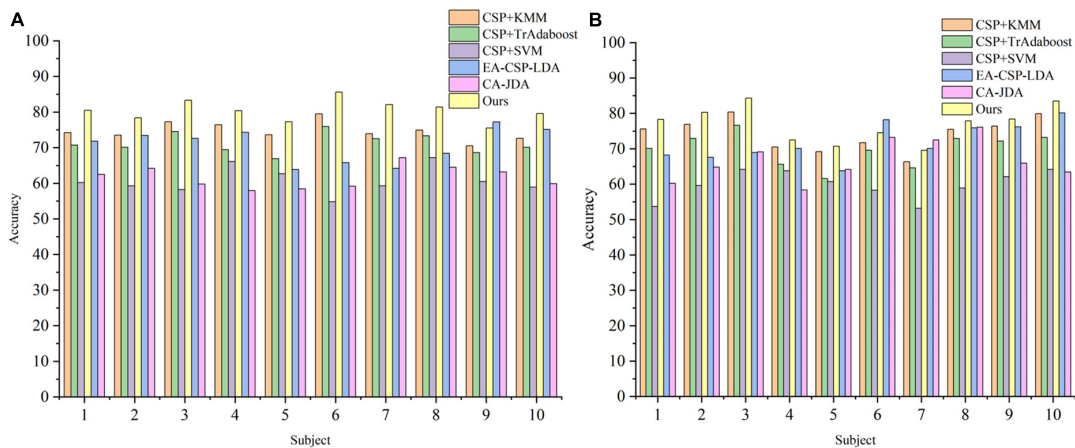
To verify the effectiveness of the proposed algorithm (Ours), ablation experiments are carried out for the following model: CSP+KMM, CSP+TrAdaboost, and CSP+SVM.

The ablation experiment is carried out on public dataset 1 and public dataset 2, and the results are shown in Table 2. It can be seen from Table 2 that the average classification accuracy of CSP+KMM algorithm is only 84.1%. The main reason is that KMM is an unsupervised algorithm and cannot effectively use the label information of source and target domain data. The average classification accuracy of CSP+TrAdaboost algorithm is only 77.9%, the main reason is that TrAdaboost is sensitive to the sample initialization weight matrix, while the traditional TrAdaboost algorithm assigns the same initialization weight to all samples. Compared with the traditional TrAdaboost algorithm, this paper uses the KMM algorithm to calculate the importance of samples as the initial sample weight matrix

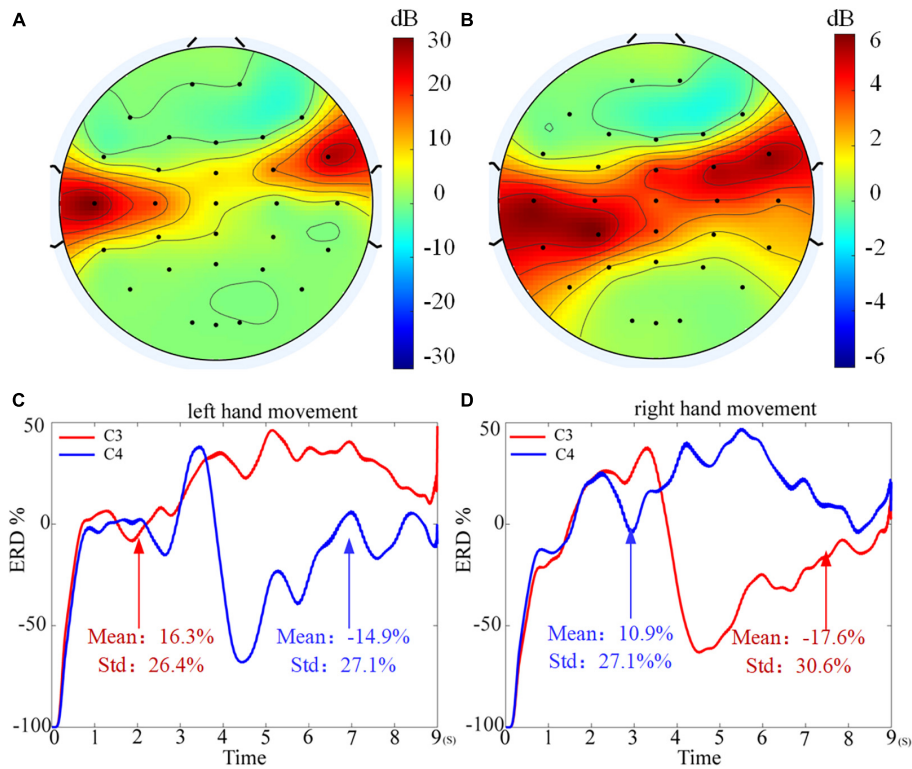
TABLE 5 Separate session of T-test statistical analysis for the proposed method vs. state-of-the-art methods.

Paired T-test	P-value	
	Ours vs. EA-CSP-LDA	Ours vs. CA-JDA
A01	*	**
A02	~	**
A03	*	*
A04	*	*
A05	~	*
A06	*	**
A07	~	**
Average	*	*

~ nonsignificant, \* $P \leq 0.05$ , \*\* $P \leq 0.01$ .



**FIGURE 7**  
The classification accuracy curve in the in-house dataset. (A) The result of table tennis group. (B) The result of non-table tennis group.



**FIGURE 8**  
Topographic map of weight distribution. (A) The result of table tennis group. (B) The result of non-table tennis group. (C) The event related desynchronization (ERD) analysis for the left hand motor imagination. (D) The ERD analysis for the right hand motor imagination.

of TrAdaboost. Experimental results verify the effectiveness of this idea, and the average classification accuracy of the proposed algorithm is 89.1%, which is higher than that of other classification methods. The average classification accuracy of the CSP+SVM algorithm is only 58.5%, which proves that there is a big difference between the data of each subject in MI task, and

the model trained by other subjects cannot be well used for the current subject.

### 3.4.3 Algorithm robustness analysis

In order to analyze the robustness of the our algorithm, different numbers of training trials were analyzed, including

five groups of 5:5, 6:4, 7:3, and 8:2, respectively. The results are shown in **Figure 6**. For different groups, the algorithm presented in this paper shows good classification accuracy and robustness, with the average accuracy of 82.0% and the highest average accuracy of 82.1% in the public dataset 1 and 2, respectively. For the groups of 5:5 and 6:4, the classification performance of the algorithm is relatively poor because there are too few training samples. In view of the above problems, we plan to carry out further research on the classification algorithm of motion imagination signals under small samples in the future. In addition, in order to measure the statistical significance of our algorithm, we conducted a *T*-test statistical analysis for the above two state-of-the-art algorithms and the algorithm in this paper. The results are shown in **Tables 3–5**.

#### 3.4.4 Application of algorithm on the in-house dataset

To verify the superiority of the proposed method, a comparative experiment is carried out on in-house dataset, and the results are shown in **Figures 7A, B**. For the local dataset, our algorithm still outperforms the other three comparison algorithms.

Meanwhile, by comparing the average classification accuracy between the table tennis group and the non-table tennis group, it can be seen that the average classification accuracy obtained by each algorithm of the table tennis group is better than that of the non-table tennis group. Previous studies have shown that when athletes repeatedly train the same skills for a long time, the number of neural synapses in the brain will change and the connections between various brain regions will be strengthened (Ericsson and Lehmann, 1997). Correspondingly, it will also enhance the activation degree of the motor center, which can help athletes to complete training more quickly and stably (Scott, 2004). In order to show whether the corresponding brain region will have corresponding responses during motor imagination, the topographic map are drawn in the **Figure 8** (see **Supplementary Appendix 1** for drawing details). For the two groups of subjects, it can be seen that the response area is at the regions corresponding to the primary motor cortex. From the color bar on the right, the non-table tennis group has a relatively weak degree of activation, while the participants who could play table tennis has a relatively strong degree of activation. Further, for the ERD analysis (see **Supplementary Appendix 2** for drawing details), **Figure 8C** shows that the right areas (C4) shows a power decrease in specific frequency bands when the left hand movement is imagined, while **Figure 8D** shows that the left areas (C3) shows a power decrease in specific frequency bands when the right hand movement is imagined. Based on the above theory, we can try to carry out rehabilitation training for stroke patients for a certain period of time, which has certain clinical guiding significance for brain rehabilitation treatment.

## 4 Conclusion

Facing the small number of MI-EEG training samples of a single subject and the large individual differences of MI-EEG signals among different subjects, an instance transfer learning algorithm based on KMM-Tradaboost was proposed. In this paper, the algorithm firstly preprocesses the source and target domain data, and then extracts the spatial features of the preprocessed data by using the CSP algorithm. After obtaining the spatial features, the KMM algorithm is used to calculate the sample weight matrix and initialize the TrAdaBoost algorithm. Finally, a strong classifier is trained by the TrAdaBoost algorithm after the initial weights are assigned, which is used to classify the MI-BCI data. The experimental results show that the algorithm proposed in this paper is superior to other algorithms, and provides a new idea to solve above problems. At the same time, this paper also applies the proposed algorithm to the in-house dataset, the results verify the effectiveness of the algorithm again, and the results of this study have certain clinical guiding significance for brain rehabilitation.

## Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

JF and YLiu designed the research. JF wrote the manuscript. CJ and ML collected the data and supervised the study. YLi and QH supervised and revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Guilin University of Aerospace Technology Foundation (XJ21KT17) and foundation for improving the scientific research ability of young and middle-aged teachers in universities in Guangxi in 2022 (2022KY0791).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.1068165/full#supplementary-material>

## References

- Ahuja, A., Sehgal, S., and Bindu, H. M. (2022). A novel approach for band selection using virtual dimensionality estimate and principal component analysis for satellite image classification. *Int. J. Intell. Inf. Technol.* 18:16.
- Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV Datasets 2a and 2b. *Front. Neurosci.* 6:39. doi: 10.3389/fnins.2012.00039
- Azab, A. M., Mihaylova, L., Ang, K. K., and Arvaneh, M. (2019). Weighted transfer learning for improving motor imagery-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1352–1359. doi: 10.1109/TNSRE.2019.2923315
- Blankertz, B., Müller, K. R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159.
- Cao, J., Hu, D., Wang, Y., Wang, J., and Lei, B. (2021). Epileptic classification with deep-transfer-learning-based feature fusion algorithm. *IEEE Trans. Cogn. Dev. Syst.* 14, 684–695. doi: 10.1109/TCDS.2021.3064228
- Chen, L., Gong, A., Ding, P., and Fu, Y. (2022). EEG signal decoding of motor imagination based on euclidean space-weighted logistic regression transfer learning. *J. Nanjing Univ. Nat. Sci.* 58, 264–274. doi: 10.13232/j.cnki.jnju.2022.02.010
- Chen, L., Li, M., Su, W., Wu, M., Hirota, K., and Pedrycz, W. (2021). Adaptive feature selection-based AdaBoost-KNN with direct optimization for dynamic emotion recognition in human-robot interaction. *IEEE Trans. Emerg. Top. Comput. Intell.* 5, 205–213. doi: 10.1109/TETCI.2019.2909930
- Dai, W., Qiang, Y., Xue, G., and Yong, Y. (2007). "Boosting for transfer learning. Machine learning," in *Proceedings of the 24th international conference (ICML 2007)* (Corvallis, OR: ACM).
- Ericsson, K. A., and Lehmann, A. C. (1997). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annu. Rev. Psychol.* 47, 273–305.
- Gertton, A., Borgwardt, K. M., Rasch, M., Scholkopf, B., and Smola, A. J. (2012). A Kernel two-sample test. *J. Mach. Learn. Res.* 13, 723–773. doi: 10.1142/S0219622012400135
- Guo, D., Rush, A. M., and Kim, Y. (2021). "Parameter-efficient transfer learning with diff pruning," in *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing* (Stroudsburg, PA: Association for Computational Linguistics), 4884–4896. doi: 10.18653/v1/2021.acl-long.378
- He, H., and Wu, D. (2019). Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 399–410. doi: 10.1109/TBME.2019.2913914
- Huang, J., Smola, A. J., Gretton, A., Borgwardt, K. M., and Scholkopf, B. (2007). Correcting sample selection bias by unlabeled data. *Adv. Neural Inf. Process. Syst.* 19, 601–608.
- Iturrate, I., Montesano, L., and Minguez, J. (2013). Task-dependent signal variations in EEG error-related potentials for brain-computer interfaces. *J. Neural Eng.* 10:026024. doi: 10.1088/1741-2560/10/2/026024
- Li, S., Liu, C. H., Su, L., Xie, B., and Wu, D. (2020). Discriminative transfer feature and label consistency for cross-domain image classification. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4842–4856. doi: 10.1109/TNNLS.2019.2958152
- Lin, P. J., Jia, T., Li, C., Li, T., Qian, C., Li, Z., et al. (2021). CNN-based prognosis of BCI rehabilitation using EEG from first session BCI training. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1936–1943. doi: 10.1109/TNSRE.2021.3112167
- Narayanan, A. M., and Bertrand, A. (2019). Analysis of miniaturization effects and channel selection strategies for EEG sensor networks with application to auditory attention detection. *IEEE Trans. Biomed. Eng.* 67, 234–244. doi: 10.1109/TBME.2019.2911728
- Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q. (2010). Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* 22, 199–210. doi: 10.1109/TNN.2010.2091281
- Saini, M., Satija, U., and Upadhyay, M. D. (2020). Wavelet based waveform distortion measures for assessment of denoised EEG quality with reference to noise-free EEG signal. *IEEE Signal Process. Lett.* 27, 1260–1264. doi: 10.1109/LSP.2020.3006417
- Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nat. Rev. Neurosci.* 5, 532–546.
- Shibanoki, T., Koizumi, Y., Yozan, B. A., and Tsuji, T. (2018). "Selection of motor imageries for brain-computer interfaces based on partial Kullback-Leibler information measure," in *Proceedings of the IEEE life sciences conference*, Montreal, QC, 243–246. doi: 10.1109/LSC.2018.8572046
- Tan, X., Liu, Y., Li, Y., Wang, P., Zeng, X., Yan, F., et al. (2018). Localized instance fusion of MRI data of Alzheimer's disease for classification based on instance transfer ensemble learning. *Biomed. Eng. Online* 17:49. doi: 10.1186/s12938-018-0489-1
- Tang, Z., Zhang, L., Chen, X., Ying, J., Wang, X., and Wang, H. (2022). Wearable supernumerary robotic limb system using a hybrid control approach based on motor imagery and object detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 1298–1309. doi: 10.1109/TNSRE.2022.3172974
- Wan, Z., Yang, R., Huang, M., Zeng, N., and Liu, X. (2021). A review on transfer learning in EEG signal analysis. *Neurocomputing* 421, 1–14. doi: 10.1016/j.neucom.2020.09.017
- Wang, H., Xu, G., Wang, X., Sun, C., Zhu, B., Fan, M., et al. (2019). The reorganization of resting-state brain networks associated with motor imagery training in chronic stroke patients. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2237–2245. doi: 10.1109/TNSRE.2019.2940980
- Wu, H., Niu, Y., Li, F., Li, Y., and Dong, M. (2019). A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification. *Front. Neurosci.* 13:1275. doi: 10.3389/fnins.2019.01275
- Xie, X., Yao, H., Gu, Z., Yu, Z., Tang, R., Xu, J., et al. (2022). Enhancement of lower limb motor imagery ability via dual-level multimodal stimulation and sparse spatial pattern decoding method. *Front. Hum. Neurosci.* 16:975410. doi: 10.3389/fnhum.2022.975410
- Xu, F., Miao, Y., Sun, Y., Guo, D., Xu, J., Wang, Y., et al. (2021). A transfer learning framework based on motor imagery rehabilitation for stroke. *Sci. Rep.* 11:19783. doi: 10.1038/S41598-021-99114-1