Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3893866, 13 pages
https://doi.org/10.1155/2022/3893866

Research Article

Motor Imagery Recognition Method Based on Multisource Transfer Learning and Multiclassifier Fusion

Chang Gao [1,2] and Jie Sun [1]

[1]*School of Information Science and Engineering, Yanshan University, Qinhuangdao, 066004 Hebei, China*
[2]*The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao, 066004 Hebei, China*

Correspondence should be addressed to Chang Gao; gc_ysu@ysu.edu.cn

Received 1 June 2022; Accepted 28 June 2022; Published 11 July 2022

Academic Editor: Chia-Huei Wu

There are two common problems in the field of motor imagery (MI) recognition, which are poor generalization and low recognition performance. A recognition method based on multisource transfer learning and multiclassifier fusion is therefore proposed to realize the MI classification. In this approach, multisource transfer learning method is used to transfer samples from multiple source domains to target domain. The source domain selection method based on distribution similarity is designed to select those source domains whose distribution is similar to the target domain, and samples with high information entropy are selected from these source domains for transferring. Then, an MI classification method is proposed through the fusion of multiple classifiers. The classifiers are trained by labeled samples in the target domain and the transferred samples in multiple source domains. The new sample in the target domain can be identified by the weight fusion of the results of these classifiers. In order to verify the effectiveness of the proposed method, four types of motor imagery in the BCI Competition IV dataset 2a were used to evaluate the recognition ability, and the results approved an excellent recognition and generalization performance as well as a better training efficiency comparing to the well-applied methods nowadays.

1. Introduction

Motor imagery (MI) has been applied to many situations, such as the recovery of patients with limb motor dysfunction and control of robotic arms, and is one of the most widely used paradigms in brain–computer interfaces (BCI) [1, 2]. It is a type of body movement (e.g., of the hands, feet, and tongue) imagined by the mind, which does not require any external stimuli. These imagined movements can be distinguished by decoding the modulation of the brain rhythms in the involved cortical areas.

As a noninvasive technique for recording brain activity, electroencephalogram (EEG) has the characteristics of high time resolution and convenient acquisition [3]; therefore, it is widely used in MI-based BCI. However, there are individual differences and time-varying owing to the different scalp shapes among human brains, and the characteristics of the human brain state change depending on time. This causes the data distribution of MI to be significantly different. Thus, it significantly affects the generalization and stability of MI recognition based on traditional machine learning methods, such as logistic regression (LR) [4, 5], support vector machine (SVM) [6, 7], linear discriminant analysis (LDA) [8, 9], and artificial neural network (ANN) [10, 11].

Transfer learning is a new machine learning method that transfers existing knowledge in the source domain to improve the poor performance of classification methods caused by small training samples in the target domain and inconsistent distribution between training and test data [3]. Therefore, transfer learning has attracted increasing attention in recent years. In MI recognition methods based on transfer learning, the target domain is a new subject with a small number of training samples, and the source domains are other subjects with a sufficient number of training samples, in which each subject can be considered as a source domain.

Knowledge distillation method that transfers the knowledge learned from a large network to a small network model to improve the classification performance of the cognitive states was proposed [12]. In [13], a general representation method for the EEG signal between different subjects or sessions based on variational autoencoders and antagonistic networks was proposed. In steady-state visual evoked potential-based BCI, Wu et al. proposed a cognitive state recognition method based on active transfer learning (ATL), by directly combining the training data of subjects in the source domain with training samples of subjects in the target domain, in the process of active learning [14]. Wu et al. combined the ATL method with a semi-supervised learning algorithm and proposed an active semi-supervised transfer learning (ASTL) method to further improve cognitive state recognition performance [15]. Hossain et al. proposed an information transfer active learning (ITAL) method based on the ATL method [16]. They investigated the information subspace in the source domain that transfers information-rich samples from the source domain to the target domain. Jeon et al. proposed an MI recognition method based on domain adaptation, using the training data of the source and target domains to train a deep neural network in an adversarial manner [17]. A multisource fusion transfer learning (MFTL) is proposed in Ref. [18]. In this method, Riemannian geometry alignment algorithm is used to select the source subjects whose features are similar to the current user. Then the balance distribution adaptation is used to calibrate the features extracted from the Riemannian tangent space. Finally, Takagi-Sugeno-Kang (TSK) fuzzy system is used to recognize the EEG signals. She et al. proposed hierarchical semi-supervised extreme learning machine (HSS-ELM) for MI recognition [19]. This method firstly employed a hierarchical ELM (H-ELM) to extract features automatically, and then the semi-supervised ELM (SS-ELM) is used to MI recognition.

In the above methods, a large number of labeled training samples are required, whereas the labeling of EEG data requires a large amount of time and economic cost. Meanwhile, these methods fail to consider large individual differences among subjects, resulting in negative transfer and severely affecting the performance of recognition methods. Further, most of these methods are aimed at binary classification, that is, left-right hand MI, which significantly hinders the practical application of BCI technology based on MI recognition.

To solve these problems, this paper proposes an MI recognition method based on multisource tranfer learning and multiclassifier fusion (MSTL-MCF). In our approach, the labeled samples in some source domains with similar data distributions as the target domain were selected based on transfer learning. Labeled samples in the source domains were obtained using a small number of labeled samples to label a large number of unlabeled samples based on semi-supervised learning. A multiclassifier fusion method was proposed to recognize the classification of an MI task. Multiple classifiers were trained using the selected labeled samples and labeled samples in the target domain. The average accuracy of cross-validation of each classifier on its own training set was considered as the weight. The final recognition result was the classification that corresponded to the maximum probability-weighted fusion of these classifiers.

The remainder of this paper is organized as follows. In Section 2, the MSTL-MCF method for MI recognition is described. In Section 3, BCI Competition IV dataset 2a is used to evaluate the performance of the proposed MSTL-MCF, including the performance of the semi-supervised learning labeling method for labeling unlabeled samples in the source domains, effectiveness of the multisource transfer learning, and recognition performance of the MI recognition based on the MSTL-MCF. Finally, Section 4 concludes the study.

2. MI Recognition Method Based on the MSTL-MCF

There are two stages in the proposed MI recognition method based on MSTL-MCF: training and recognition. A schema of the training and recognition processes is shown in Figure 1.

In the training process of MSTL-MCF, the EEG samples were first preprocessed, including filtering and eliminating artifacts. A sample alignment method in the source domains based on Euclidean space alignment (EA) was proposed to solve the problem of data distribution differences owing to the time-varying characteristics of EEG signals. Then, a large number of unlabeled samples from each source domain were labeled by the semi-supervised learning labeling method based on cotraining after extracting features using a common spatial pattern (CSP).

The following is an iterative process, including the sample transfer method based on multisource transfer learning and training of the multiclassifier fusion model. In the former method, the distribution similarity between the labeled samples of the target domain and that of each source domain was measured. M source domains with a higher distribution similarity were selected from N source domains. The labeled samples with larger information entropy in these source domains were transferred, and M training pools were formed with the labeled samples from the target domain to train the $M + 1$ classifiers. Moreover, a multiclassifier fusion model was combined with these classifiers using the classifier trained by the labeled samples in the target domain. Unlabeled samples in the target domain were classified by this fusion model, and unlabeled samples with low confidence were labeled based on active learning and added to the labeled sample set of the target domain. These steps were repeated until the specified maximum number of iterations was reached or no samples were labeled.

The new sample of the target domain was recognized by each classifier of the multiclassifier fusion model, and the results of each classifier were fused based on the weight fusion method as the final recognition result.

2.1. Sample Alignment Method Based on EA. Owing to the time-varying characteristics, the data distribution of the EEG was different with the change in time, even if the same subject performs the same MI task. Specifically for data in

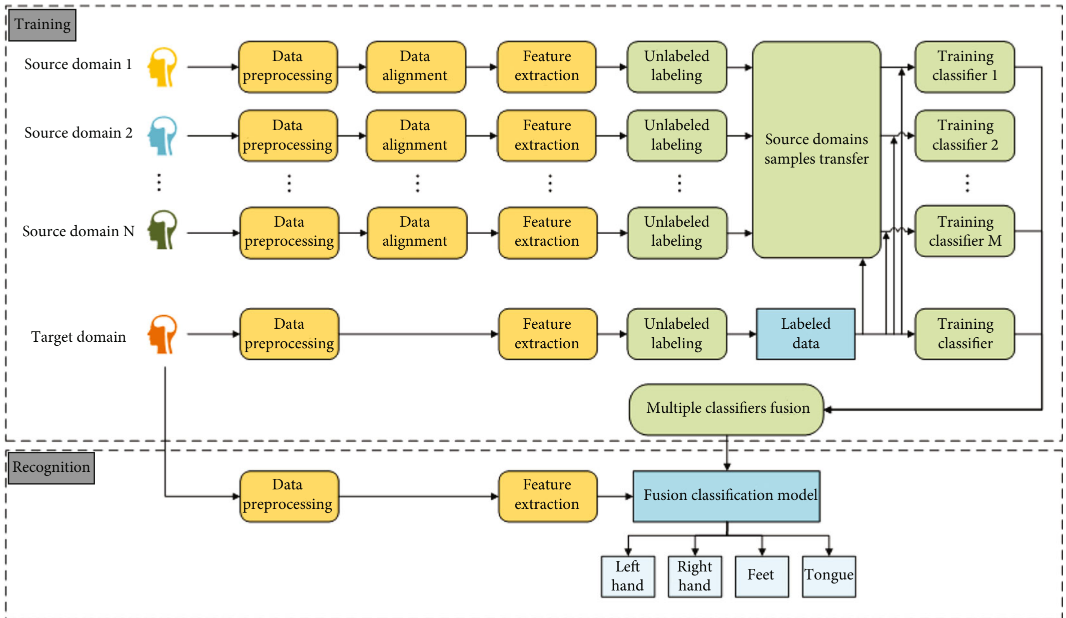

FIGURE 1: The schema of MI recognition method based on the MSTL-MCF.

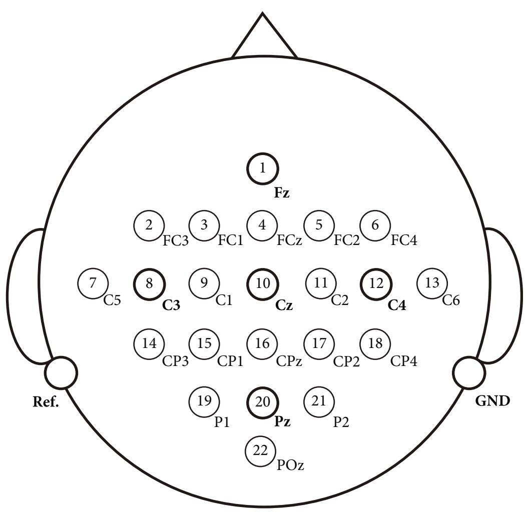

FIGURE 2: 22 Ag/AgCI electrodes [29]

the source domain with a large number of samples that are typically collected from different time periods, it is more susceptible to time-varying effects.

This paper proposes a source-domain sample alignment method based on EA. By adjusting the EEG covariance matrix, we find a projection matrix that minimizes the distance between the average covariance matrices of the same subject in different time periods to solve the time-varying influence.

For more details about EA, please refer to Ref. [20].

2.2. Feature Extraction Method. In MSTL-MCF, a feature extraction method for labeled samples of all source and target domains was proposed based on CSP [21].

First, the covariance matrix of the samples of the jth class is calculated as

$$P_j = \frac{1}{N_j}\sum_{i=1}^{N_j} S_{(j,i)} S_{(j,i)}^{\mathrm{T}}, \tag{1}$$

where $P_j (j = 1, \cdots, K, K$ is the number of classes) is the covariance matrix, T is the transpose of a matrix, $S_{(j,i)} \in \mathbb{R}^{n\times t}$ (n is the number of channels, and t is the number of time samples) is the ith sample of the jth class, and N_j is the number of labeled samples in the jth class.

Then, the spatial filter W_j of the jth class was constructed to satisfy the following equation:

$$P_j W_j = \left(C_j + C_{-j}\right) W_j D_j, \tag{2}$$

where P_{-j} is the covariance matrix of all samples except the jth class and D_j represents a diagonal matrix consisting of eigenvalues of P_j. All samples of the jth class were filtered by W_j to obtain the spatial filter signals $Z_{(j,i)} (i = 1, \cdots, N)$, where N is the total number of samples, as expressed below.

$$Z_{(j,i)} = S_i^{\mathrm{T}} W_j. \tag{3}$$

Furthermore, the m front and m back rows of $Z_{(j,i)}$ are obtained to form a new matrix $Z^{\sim}_{(j,i)}$, and the eigenvector $x_{(j,i)}$ of the ith labeled sample of the jth class is calculated as follows:

$$x_{(j,i)} = \log\left(\mathrm{var}\left(\tilde{Z}_{(j,i)}\right)\right). \tag{4}$$

Table 1: Comparison of the accuracy of different MI recognition methods (%).

Subjects	Method MFTL	DA_ PSD	HSS- ELM	ITAL_ CSP	ITAL_ FBCSP	MSTL_ MCF
1	79.00	70.00	81.14	67.50	81.30	82.57
2	52.00	38.00	49.86	46.80	49.50	53.89
3	85.75	76.00	78.02	74.90	84.65	86.50
4	58.75	39.00	63.33	28.60	58.00	67.60
5	45.25	31.00	44.03	30.50	50.90	43.67
6	49.00	38.00	49.44	45.90	45.50	48.58
7	82.00	70.00	81.11	75.00	82.30	86.13
8	82.00	63.00	81.49	77.40	81.60	84.59
9	80.50	59.00	81.38	79.40	72.50	84.23
Mean	68.50	53.78	67.76	58.70	67.10	70.86

Finally, the above steps are iterated K times to obtain the eigenvector of the ith sample $x_i = [x_{(1,i)}^T, x_{(2,i)}^T, \cdots, x_{(K,i)}^T]^T$, which is a $K \times 2m$ dimension column vector.

In the feature extraction of unlabeled samples, the covariance matrix and spatial filter were estimated from labeled samples in the same domain.

2.3. Sample Labeling Based on Semisupervised Learning. In the proposed MSTL-MCF, many labeled samples are necessary in each source domain. However, manual labeling requires much time and economic costs. We propose a semi-supervised learning labeling method for unlabeled samples of the source domains.

2.3.1. Semi-supervised Learning Labeling Based on Cotraining. To increase the labeling accuracy, we introduced cotraining into the unlabeled sample labeling method based on semi-supervised learning. The process was as follows:

(1) The labeled sample sets $X^{(L_1)}$ and $X^{(L_2)}(X^{(L_1)} = X^{(L_2)})$ were used to train two classifiers: SVM (C_{SVM}) and LR (C_{LR}).

(2) The samples in the unlabeled sample sets $X^{(U_1)}$ and $X^{(U_2)}$ ($X^{(U_1)} = X^{(U_2)}$) were labeled by the trained classifiers C_{LR} and C_{SVM}, respectively, and pseudolabeled sample sets $X^{(PL_1)}$ and $X^{(PL_2)}$ were obtained

(3) The confidence of the samples in $X^{(PL_1)}$ and $X^{(PL_2)}$ was measured, and it was determined whether the labels of the samples with high confidence were consistent with the labels of $K(K = 3)$, labeled samples adjacent to them. If so, these pseudolabeled samples were added to the labeled sample sets $X^{(L_1)}$ and $X^{(L_2)}$, and the C_{SVM} and C_{LR} classifiers were retrained

Using the retrained classifiers C_{SVM} and C_{LR}, label the unlabeled sample sets $X^{(U_2)}$ and $X^{(U_1)}$, respectively, and repeat the process of the last paragraph until the maximum number of iterations is reached or when no sample in the unlabeled sample sets can be added to the labeled sample sets. The intersection of the two labeled sample sets obtained by the last iteration is taken as the final labeled sample set; that is, $X^{(L)} = X^{(L_1)} \cap X^{(L_2)}$.

Furthermore, during this process, the labeling of unlabeled samples may change as the number of iterations increases. To avoid samples with uncertain categories residing in the labeled sample sets, a dynamic adjustment mechanism for labeled samples was proposed. Consequently, the labeling the accuracy of the unlabeled samples in the source domains was improved. Additionally, it provided more transferable samples with reliable category information.

2.3.2. Sample Confidence Measurement. Mislabeled samples lead to a decrease in the labeling accuracy of unlabeled samples if they participate in the training of the classifier. Therefore, in the MSTL-MCF, we propose the concept of a confidence measure and estimate the possibility of correct labeling of the samples. The computational formula is given by

$$M_i = \frac{\omega_i}{E_i}, \tag{5}$$

where M_i, ω_i, and E_i are the confidence measure, reliability coefficient, and information entropy of the ith pseudolabeled sample, respectively. The greater the confidence level of the sample, the more likely it is to belong to this category. Selecting samples with higher confidence to add to the labeled sample sets can avoid the impact of mislabeled samples on training the classifiers.

The reliability coefficient of the sample indicates how reliable it belongs to its labeled category, and the computational formula is given as follows:

$$\omega_i = \frac{P_i^{(1)} - P_i^{(2)}}{P_i^{(1)}}, \tag{6}$$

where $P_i^{(1)}$ represents the posterior probability of the ith unlabeled sample of the classification that is labeled and $P_i^{(2)}$ represents one of the ith unlabeled samples of the slave classification with the second largest posterior probability value. The larger the value of ω_i, the greater the reliability of the ith unlabeled sample belonging to the corresponding classification.

The information entropy is used to measure the amount of information in a sample [22–24]. The smaller the information entropy, the less category information the samples contain, and vice versa. The formulation is as follows:

$$E_i = -\sum_{c=1}^{n_c} P(y_c|x_i) log P(y_c|x_i), \tag{7}$$

where $P(y_c|x_i)$ represents the posteriori probability that x_i belongs to the cth class and n_c represents the number of classifications.

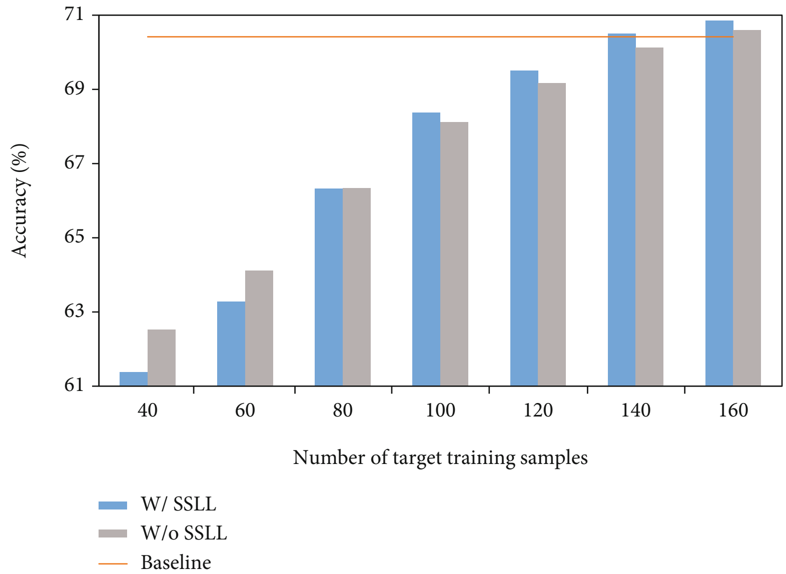

Figure 3: The recognition performance of w/ SSLL, w/o SSLL and baseline.

Table 2: Labeling accuracy of unlabeled samples in multiple source domains based on semi-supervised learning(%).

Subjects	The number of correct labeled samples	The number of mislabeled samples	Accuracy(%)
1	127	10	92.70
2	144	28	83.72
3	184	39	82.51
4	302	34	89.88
5	344	56	86.00
6	508	71	87.74
7	262	25	91.29
8	207	33	86.25
9	254	31	89.12

2.4. Sample Transfer Based on Multisource Transfer Learning. When the number of samples in the target domain is small, it is difficult to obtain a classification model that can accurately identify the MI classification of the samples in the target domain. Moreover, it takes a considerable amount of time to collect sufficient training data for the target domain. Therefore, it is necessary to transfer a number of labeled samples from the source to target domains. However, if the samples of the source domains whose data distribution is significantly different from that of the target domain, the phenomenon of negative transfer occurs, resulting in a significant decline in the MI recognition performance of the target domain.

In MSTL-MCF, we propose a method for transferring samples based on multisource transfer learning, in which a large number of labeled samples in some source domains are transferred to the target domain. The proposed method consists of two stages. First, by measuring the data distribution similarity between the target domain and each source domain, source domains with large data distribution similarities are selected. Second, samples with high information entropy from these source domains are selected to improve the robustness of our approach.

2.4.1. Source Domain Selection. In the source domain selection method, we propose a distribution similarity measure method between the target domain and each source domain. The formulation is as follows:

$$B_k = \frac{1}{\left\| \sum_{i=1}^{n^{(T)}} \phi\left(x_i^{(T)}\right) - \sum_{j=1}^{n^{(S_k)}} \phi\left(y_j^{(S_k)}\right) \right\|_H^2}, \quad (8)$$

where B_k is the distribution similarity of the samples between the target and the kth source domains. $\| \sum_{i=1}^{n^{(T)}} \phi(x_i^{(T)}) - \sum_{j=1}^{n^{(S_k)}} \phi(y_j^{(S_k)}) \|_H^2$ is called maximum mean discrepancy [25], which is used to measure the difference between two different but related distributions. Samples of the source domain and that of the target domain are, respectively, mapped to the Reproducing Kernel Hilbert Space by a kernel function, i.e., $\phi(\bullet)$, so as to avoid the problem of infinite MMD distance between the two distributions. $n^{(T)}$ and $n^{(S_k)}$ are the number of samples in the target domain and the kth source domain, respectively; $x_i^{(T)}$ is the feature vector of the ith sample of the labeled sample set $X^{(T)}$ in the target domain; and $y_j^{(S_k)}$ is the feature vector of the jth sample in the kth source domain.

A simple threshold B_{th} is defined to guarantee a high estimation precision for transferring suitable source domains. If $B_k > B_{th}$, the data distribution between the target domain and the kth source domain is similar, and the samples in this source domain can be transferred. When B_{th} is set between 3.0 and 4.5, the difference between the highest

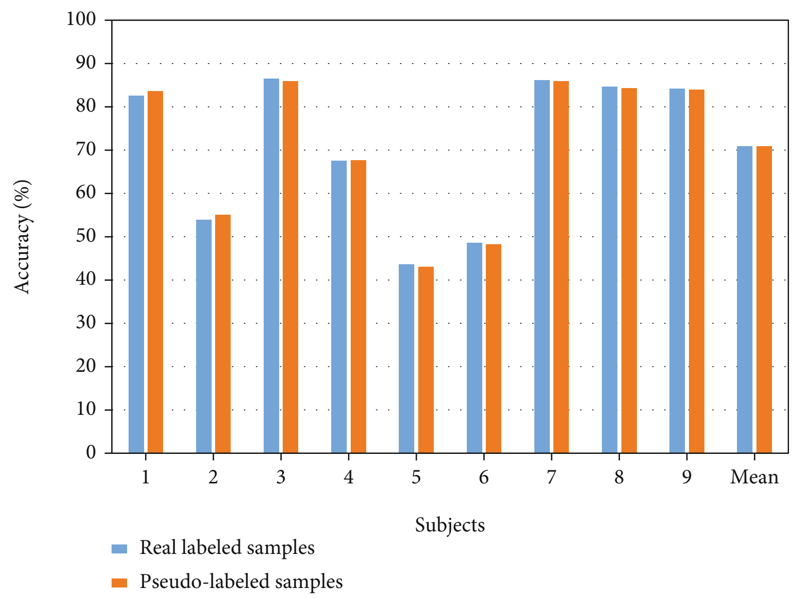

Figure 4: The accuracy of the recognition models trained by real labeled samples and pseudo-labeled samples.

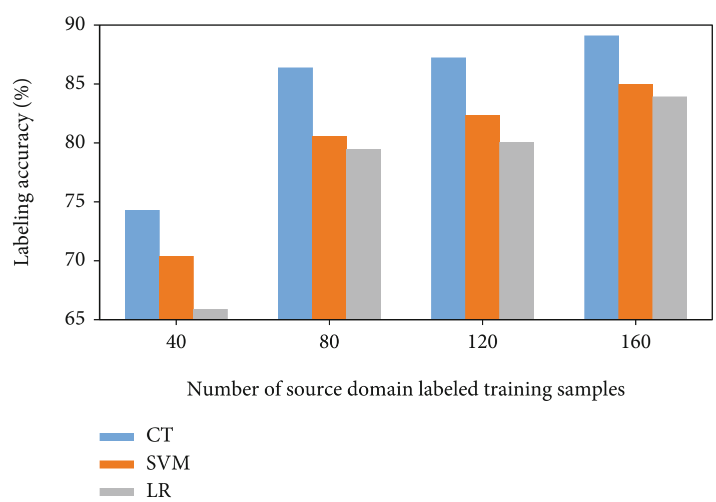

Figure 5: The labeling accuracy of CT, SVM and LR.

($B_{th} = 3.5$) and the lowest accuracy for motor imagery recognition is about 0.03%. Therefore, the concrete value of the similarity threshold is set to 3.5.

M source domains with similar data distribution to the target domain are selected from all N source domains. The specific number of M depends on how many source domains have B_k greater than B_{th}. The target domain is different, and the specific value of M is also different.

2.4.2. Unlabeled Sample in the Target Domain Labeling Method Based on Active Learning. In the proposed MSTL-MCF, labeled samples of the target domain are obtained based on active learning. In active learning, some labeling requests can actively generate and submit selected samples to experts for labeling [26, 27]. In our approach, the expert labeling results are simulated using the real label of the sample. In each iteration, the unlabeled samples in the target domain are predicted by the fusion classification model.

First, the unlabeled samples in the target domain are predicted using the fusion classification model.

Then, an active query method based on sample uncertainty is proposed in which the sample confidence calculated based on (5) is used as the sample selection criterion, and samples with low sample confidence are selected for labeling to form a newly labeled sample set $X^{(T)}$.

Finally, classifier C_0 is retrained by $X^{(T)}$, and it is further used to reselect the transferable source domains and samples from all source domains.

2.4.3. Source Domain Sample Transfer. The labeled samples in the distribution similarity source domains are classified using classifier C_0. If the classification results are consistent

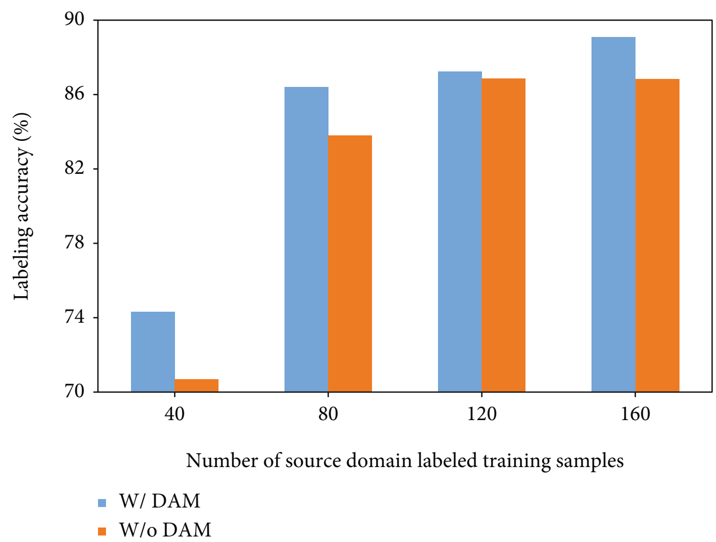

Figure 6: The labeling accuracy of w/ DAM and w/o DAM.

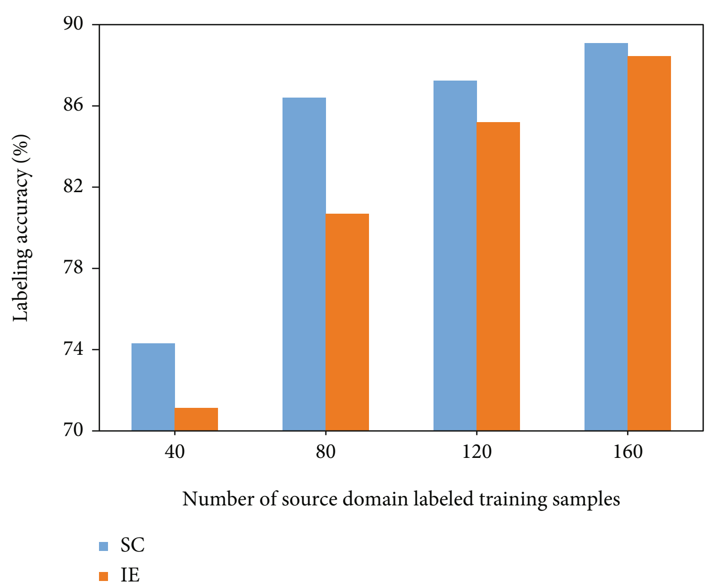

Figure 7: The labeling accuracy of SC and IE.

with the labeled information, the data distribution of these samples is similar to that in the target domain.

For these samples, the information entropy $E_{(j,i)}$ (the ith sample in the jth similarity source domain) is calculated according to (7). There is an information entropy threshold E_{th}. If $E_{(j,i)} > E_{th}$, it indicates that the sample is uncertain and has a large amount of information; then, the sample is selected to enhance the robustness of the recognition method. The selected samples comprise the labeled sample sets $X_j^{(S)} (j = 1, 2, \cdots, M)$ in each similar source domain. It was also found in the experiments that when E_{th} is set between 0.7 and 0.9, recognition accuracy of MI tasks was hardly influenced. The threshold E_{th} is defined to 0.8, to guarantee high detection precision for suitable sample transfer.

2.5. MI Recognition Based on Multiclassifier Fusion. To solve the problem of insufficient training of the classification model caused by the lack of training samples in the target domain, we propose a multiclassifier fusion method for MI recognition of the target domain. The multiclassifier fusion model comprises $M + 1$ classifiers, $C_0, C_1, \cdots, C_M$, and each classification is assigned a weight. Based on the concept of weight fusion, the results of the sample to be recognized by each classifier are fused as the final recognition result. This study implements a multiclassifier fusion model on multiple logistic regression (LR) classifiers. The number of classifiers in this model depends on how many source domains have a similar data distribution to the target domain.

First, the target domain classifier C_0 is trained by the labeled sample set $X^{(T)}$ in the target domain, and the classifier

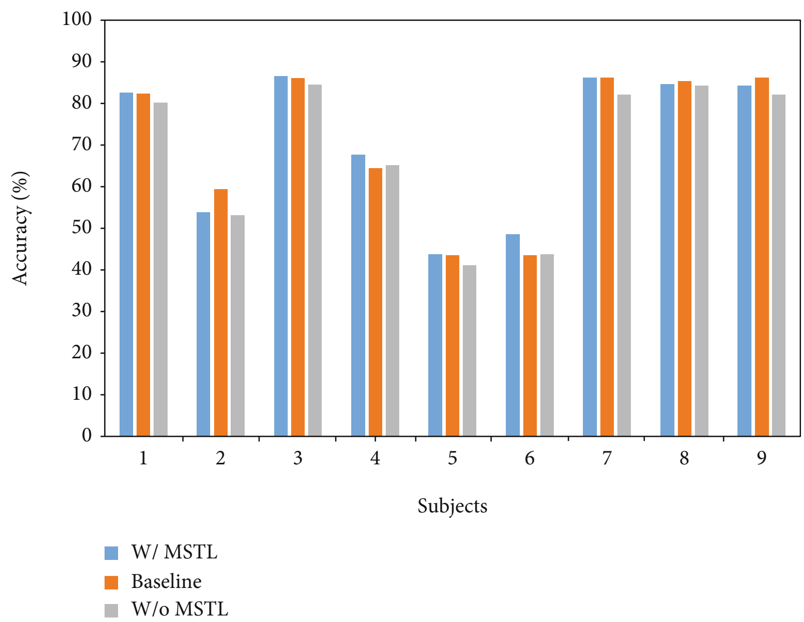

Figure 8: The recognition accuracy of the 9 subjects, respectively, as the target domain.

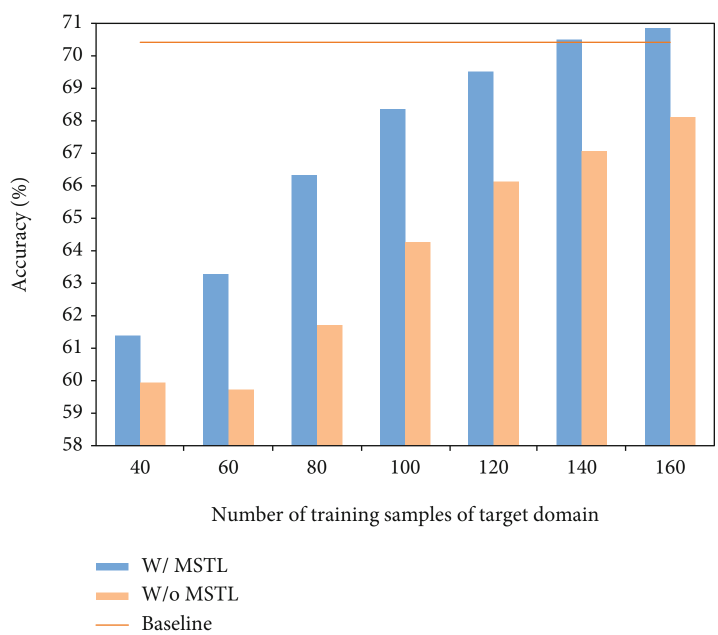

Figure 9: The average recognition accuracy.

parameter $\omega^{(T)}$ is obtained by optimizing the objective function $J_{\omega^{(T)}}(\bullet)$:

$$J_{\omega^{(T)}} = \min_{\omega^{(T)}} \left[-\frac{1}{n^{(T)}} \sum_{i=1}^{n^{(T)}} \left(\begin{array}{c} y_i^{(T)} \log h_{\omega^{(T)}}\left(x_i^{(T)}\right) + \\ \left(1 - y_i^{(T)}\right) \log \left(1 - h_{\omega^{(T)}}\left(x_i^{(T)}\right)\right) \end{array} \right) + \lambda^{(T)} \left\| \omega^{(T)} \right\|_2^2 \right], \quad (9)$$

where $n^{(T)}$ is the number of labeled samples in the target domain; $x_i^{(T)}$ and $y_i^{(T)}$ are the feature vector and label of the i th labeled sample in the target domain, respectively; $h_{\omega^{(T)}}(\bullet)$ is the discriminant function of LR, which is a sigmoid function [28]; and $\lambda^{(T)}$ is the regularization parameter.

Then, classifiers $C_j^{(S)}(j = 1, 2, \cdots, M)$ are trained by $X^{\sim}{}_j^{(S)}(j = 1, 2, \cdots, M)$, and the parameter $\omega_j^{(S)}(j = 1, 2, \cdots, M)$ of each classifier is obtained by optimizing the objective

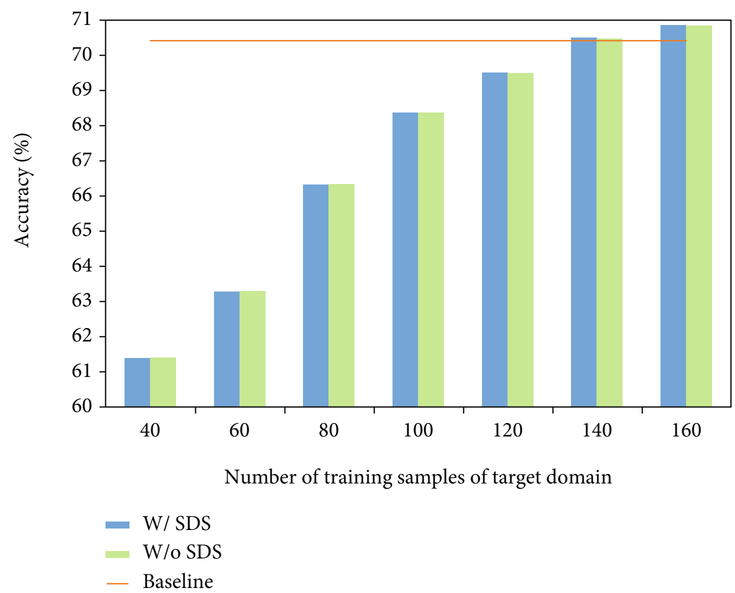

Figure 10: The average recognition accuracy of w/ SDS and w/o SDS.

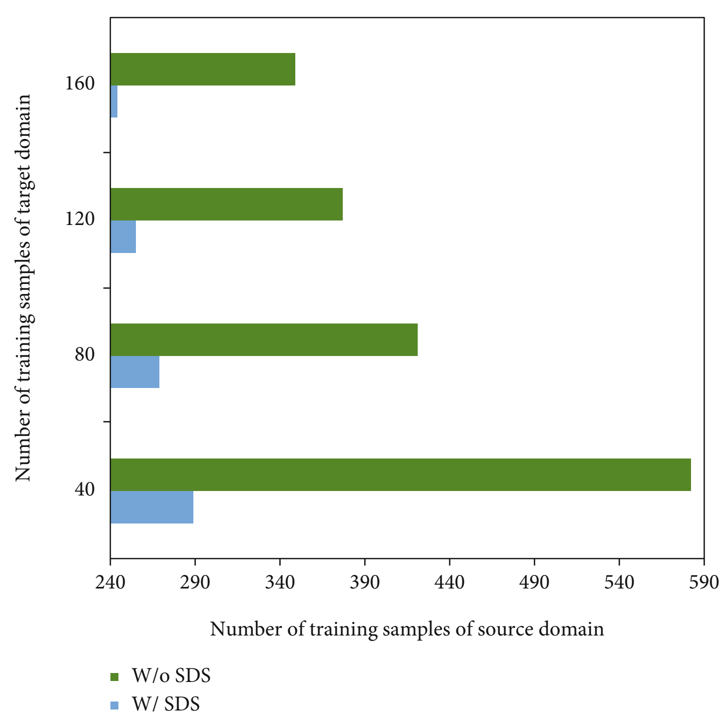

Figure 11: The average number of the source domain samples transferred by the w/ SDS method and w/o SDS method.

function $J_{\omega_j^{(S)}}(\bullet)$, respectively:

$$J_{\omega_j^{(S)}} = \min_{\omega_j^{(S)}} \left[-\frac{1}{n^{(T)} + \tilde{n}_j^{(S)}} \sum_{i=1}^{n^{(T)}+\tilde{n}} \left(\begin{array}{c} \tilde{y}_{(j,i)}^{(S)} \log h_{\omega_j^{(S)}}\left(\tilde{x}_{(j,i)}^{(S)}\right) + \\ \left(1 - \tilde{y}_{(j,i)}^{(S)}\right) \log \left(1 - h_{\omega_j^{(S)}}\left(\tilde{x}_{(j,i)}^{(S)}\right)\right) \end{array} \right) + \lambda_j^{(S)} \left\| \omega_j^{(S)} \right\|_2^2 \right], \tag{10}$$

where $n^{\sim}{}_j^{(S)}$ is the number of samples in the $X^{\sim}{}_j^{(S)}$, and $x^{\sim}{}_{(j,i)}^{(S)}$ and $y^{\sim}{}_{(j,i)}^{(S)}$ are the feature vector and label of the ith labeled sample in the $X^{\sim}{}_j^{(S)}$, respectively.

Finally, each classifier of the $C_j (j = 1, 2, \cdots, M)$ performs 10-fold cross-validation on its own training set, and its average accuracy $\alpha_j (j = 1, 2, \cdots, M)$ is taken as the weight of each classifier. The weight of C_0 is one because it is directly related to the samples in the target domain.

The result of the last iteration is used as the final fusion classification model. When a new sample is recognized, it is predicted by $M + 1$ classifiers and the probability of the sample belonging to a certain class in each classifier is weighted to obtain the probability that the sample belongs to the class. The corresponding category when the probability is the maximum is the classification to which the sample to be identified:

$$l = \underset{k}{argmax}\left(p_k^{(C_0)} + \sum_{j=1}^{M} \alpha_j \times p_k^{(C_j)}\right), \tag{11}$$

where l indicates the classification that the sample is identified and $p_k^{(C_0)}$ and $p_k^{(C_j)}$ are the probabilities that the sample to be identified is the kth classification of the C_0 and C_j.

The recognition method based on the fusion of multiple classifiers can avoid insufficient training of a single classifier owing to the lack of training samples, which affects the recognition performance. By assigning different weights to each classifier, the influence of low-performance classifiers on the entire classification model can be reduced, thereby improving the accuracy of the model.

3. Experimental Results and Effectiveness Evaluation

To evaluate the performance of our approach, we verified the effectiveness of the MI recognition method based on MSTL-MCF and analyzed the results. To obtain general results, all results were averaged after 30 cycles.

3.1. Experimental Setup. Proposed method was applied to BCI Competition IV dataset 2a [29]. As shown in Figure 2, the dataset consists of 22 EEG channels collected from 9 healthy subjects with a sampling frequency of 250 Hz, including Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, and POz. Raw EEG signal was filtered by a third-order Butterworth band-pass filter of 8-32 Hz. Among them, each subject recorded 576 trails involving four MI tasks, including imagination of movement of the left hand, right hand, both feet, and tongue, of which includes two sessions with 72 trails for each class. In the following experiment, the time segment of 0-4 s after the onset of the visual cue was used as one trail data, that is, one sample. In order to reduce the influence of the electrooculogram (EOG) activity on the EEG, regression analysis is used in this study. It is the most challenging dataset for MI classification and the most commonly used published dataset for evaluating the performance of MI classification methods, which has a low signal-to-noise ratio, significant differences between subjects, and more MI tasks.

Compared with other datasets, it has considerable signal noise and extreme values,

In each source domain, samples of two sessions were combined, 80 samples (20 samples in each class) were randomly selected as the labeled samples, and others were used as the unlabeled samples. The maximum times of iteration is 20, and 20 samples were selected in each iteration.

In the target domain, samples in the one session were used as the training set, and other samples were used as the test set. 40 samples were randomly selected from the training set as the labeled samples, and others were used as unlabeled samples. The maximum times of iteration is 30, and 4 samples were selected in each iteration.

The baseline method is LR that has better recognition accuracy than SVM and LDA. All samples in one session are training set and all samples in other session are test set for each subject.

3.2. Performance Evaluation of MI Recognition Based on MSTL-MCF. For the performance evaluation of MSTL-MCF in MI recognition, the recognition results of the target domain samples were compared with those of MFTL, DA_PSD, HSS-ELM, ITAL_CSP, and ITAL_FBCSP, which were proposed in recent years and are commonly used for comparison. In the experiment, nine subjects were selected individually as the target domain, and the remaining eight subjects were used as the source domains. In these three methods, one session was used as the training set and the other as test set in the target domain.

The results are listed in Table 1.

Clearly from Table 1, the average recognition accuracy of the nine subjects of our method can reach 70.86%, which is significantly higher than that of the other methods, in which our proposed method has the best performance for all subjects, except for subjects 5 and 6. Only 160 labeled samples of the target domain were used in our method, ITAL_CSP and ITAL_FBCSP; however, all the labeled training samples of the target domain were used in the other three methods.

Therefore, our proposed method has higher recognition accuracy and training efficiency and is more conducive to the practical application of BCI technology. This is mainly because the labeling accuracy of unlabeled samples was improved through the semisupervised learning labeling method for the unlabeled samples in the source domains. The multisource transfer learning method provides a large number of labeled samples with reliable category information for the target domain. By measuring the similarity of the data distribution between the target domain and each source domain, the impact of negative transfer on the recognition performance is avoided. Finally, in the fusion recognition model of multiple classifiers, the weights of the classifiers were different, thereby reducing the impact of relatively low-performance classifiers on the fusion model to improve the performance of MI recognition.

3.3. Reliability Evaluation of the Semi-supervised Learning Labeling Method. To verify the reliability of the semi-supervised learning labeling method in MSTL-MCF, we first evaluated the influence of the proposed method on the MI recognition performance of the target domain. Subsequently, we evaluated the reliability of the method and influence of mislabeled samples in the source domains on the recognition performance of the target domain samples. Moreover, the effects of cotraining, dynamic adjustment mechanism, and sample confidence measurement on labeling performance were verified.

Data from the two sessions of each source domain were combined. Forty samples were randomly selected as the test set, and the remaining samples were used as the training set. In particular, 40, 80, 120, and 160 samples (10, 20, 30, and 40 samples in each class) were randomly selected from the training set as the initial labeled sample training set, and the remaining samples were selected as the unlabeled sample training set.

3.3.1. Labeling Effect of Unlabeled Samples in Source Domains. To verify the impact of the source domain sample labeling method based on semi-supervised learning on the recognition performance of target domain samples in the MSTL-MCF, the accuracy of MI recognition in the target domain obtained by the MSTL-MCF was compared to that of the semi-supervised learning labeling method (w/SSLL), which was not used in the proposed method (w/o SSLL). The results are shown in Figure 3. Both methods have the average results of nine subjects as the target domain.

In Figure 3, the horizontal and vertical axes represent the number of labeled training samples and MI recognition accuracy in the target domain, respectively. Clearly from the results, when the training samples in the target domain were less than 80, the recognition accuracy of the w/ SSLL was slightly lower than that of the w/o SSLL. However, with an increase in the number of training samples in the target domain, the recognition accuracy of these two methods is almost the same.

This is because when the target domain has few training samples, some of the mislabeled samples are transferred, thus affecting the recognition performance of the classification model.

3.3.2. Influence of Mislabeled Samples in the Source Domains on the MI Recognition. The proposed semi-supervised learning labeling method inevitably contains mislabeled samples. Thus, we first evaluated the influence of these mislabeled samples on the training process of the MSTL-MCF in this experiment. When every subject was taken as the target domain, we counted the number of correctly labeled samples, number of incorrectly labeled samples transferred from the source domains, and corresponding labeling accuracy of these samples. The results are listed in Table 2, where the number of labeled samples was 160 in the target domain.

Clearly from the results, the labeling accuracy is from 82.51% to 92.70%, and the average accuracy is 87.60%. Therefore, the semi-supervised learning labeling method in MSTL-MCF can supply a large number of reliable correctly labeled samples.

In the MSTL-MCF-based MI recognition method, in the two cases of using pseudolabels (including incorrectly labeled samples) to participate in the training and using real labels (all correctly labeled samples) to participate in the training, the recognition accuracies were compared. The results are presented in Figure 4.

From the results in Figure 4, the recognition accuracies were all significantly close, regardless of the accuracy of each subject or average accuracy. Therefore, in the proposed MSTL-MCF method, in the process of recognition model training, the influence on the MI recognition effectiveness of the target domain samples is significantly small, even if there are a small number of mislabeled samples in source domains.

3.3.3. Cotraining Effectiveness Verification. In the cotraining (CT) effectiveness verification experiment, the labeling effect of the cotraining of SVM and LR was compared with that of using SVM and LR, and results are shown in Figure 5.

As shown in Figure 5, the CT has a better labeling effect than SVM and LR. For different numbers of labeled samples, the labeling accuracy of the CT method was 3.92%, 5.83%, 4.86%, and 4.11% higher than that of SVM and 8.44%, 6.95%, 7.17%, and 5.15% higher than that of LR. Therefore, for the proposed semi-supervised learning labeling method in MSTL-MCF, the CT method using two classifiers for training can effectively improve the accuracy of the unlabeled samples of the source domains.

3.3.4. Effectiveness Verification of the Dynamic Adjustment Mechanism. To verify the impact of the dynamic adjustment mechanism on the effect of the semi-supervised learning labeling method, the labeling effect of unlabeled samples of the source domains was compared using the dynamic adjustment mechanism (w/ DAM) and not using the dynamic adjustment mechanism (w/o DAM). The results are shown in Figure 6.

Clearly from Figure 6, the w/ DAM method has a better labeling effect than the w/o DAM method, and the labeling accuracy improved by 3.65%, 2.58%, 0.39%, and 2.27%. Therefore, the dynamic adjustment mechanism can effectively improve the accuracy of unlabeled samples in the source domains.

3.3.5. Effectiveness Verification of the Sample Confidence Measurement. In the experiment, we used the sample confidence (SC) and information entropy (IE) methods to select the samples from the source domains to verify the effectiveness of the SC method. The labeling effectiveness of unlabeled samples in the source domains was compared. The results are shown in Figure 7.

As shown in Figure 7, for different numbers of labeled training samples in the source domains, using the SC method for sample selection has better labeling effectiveness than using the IE method, and the labeling accuracy improved by 3.21%, 5.75%, 2.05%, and 0.63%. Therefore, the sample confidence measurement method can effectively improve the labeling accuracy of unlabeled samples in the source domains.

3.4. Effectiveness Evaluation of Multisource Transfer Learning. To verify the effectiveness of the multisource transfer learning method in MSTL-MCF, the recognition performance of the target domain samples was compared using the multisource transfer learning method (w/ MSTL) and not using the multisource transfer learning method (w/o MSTL). The recognition model was trained by the labeled samples in the target domain. The recognition accuracy of the nine subjects in the target domain is shown in

Figure 8, and the average recognition accuracy is shown in Figure 9.

As shown in Figures 8 and 9, the w/ MSTL method has significantly better recognition performance than the w/o MSTL method, which shows that selecting the samples from the source domains based on multisource transfer learning can improve the MI recognition performance of the target domain.

Therefore, transferring suitable samples to participate in the training of the MI recognition model of the target domain has better recognition performance than the classification model trained only by the labeled training samples of the target domain. Furthermore, it can effectively reduce the number of labeled training samples in the target domain, thereby reducing the time required to calibrate the classification model for MI recognition.

3.5. Effectiveness Verification of the Source Domain Selection. To evaluate the effectiveness of the source domain selection (SDS) in the multisource transfer learning method, the MI recognition performance of the target domain was compared using the SDS (w/ SDS) and not using the SDS (w/o SDS). In the w/o SDS method, all labeled samples of the source domains were considered as candidate samples. The average recognition accuracy is shown in Figure 10.

As shown in Figure 10, regardless of the number of labeled samples used in the target domain to participate in the model training, the results of w/ SDS method and that of w/o SDS method were significantly close. This shows that the proposed source domain selection method can effectively determine source domains with data distributions similar to the target domain. With comparable recognition accuracy, classification model training in the target domain requires less samples in the source domain, which can improve the efficiency of model training.

To show the performance of the source domain selection method in improving the training efficiency of the classification model for the target domain more clearly, the average number of source domain samples transferred by the w/ SDS method was compared to that of the w/o SDS method when nine subjects were used as the target domain, and the results are shown in Figure 11.

The results show that the number of labeled training samples transferred from the source domains in the w/ SDS method is lower than that of the w/o SDS method, which can reduce the training time of the classification model. However, the recognition accuracy of the target domain was equivalent. Therefore, the w/ SDS method can improve the training efficiency of the classification model in the MSTL-MCF.

4. Conclusion

This paper proposes a MI recognition method based on multisource transfer learning and multiclassifier fusion to solve the problem of poor generalization of MI recognition methods due to individual differences, which is common in MI recognition methods. In our method, a semi-supervised learning labeling method is proposed to provide a large number of samples with reliable labeling which are provided for the subsequent process. In the multisource transfer learning, the source domains with high distribution similarity are selected by measuring the distribution similarity between the target domain and each source domain, and the samples with high information entropy from these source domains are transferred. In the multiclassifier fusion, several classifiers are trained by the samples in the transferred source domains and the samples in the target domain, and the weights are assigned to each classifier. Furthermore, the final recognition result is obtained based on the idea of weight fusion. Our method has an average recognition accuracy of 70.86% for the four types of motor imagery in the BCI Competition IV dataset 2a, which is approximately 3.7% higher than that of the closest method. In addition, the effectiveness of the source domain unlabeled sample labeling method and the effectiveness of the multisource transfer learning method proposed in this paper are also evaluated.

Data Availability

The [BCI Competition IV dataset 2a] data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the Innovation Capability Improvement Plan Project of Hebei Province (22567626H).

References

[1] S. Machado, F. Araújo, F. Paes et al., “EEG-based brain-computer interfaces: an overview of basic concepts and clinical applications in neurorehabilitation,” *Reviews in the Neurosciences*, vol. 21, no. 6, pp. 451–468, 2010.

[2] L. Mercado, L. Alvarado, G. Quiroz-Compean et al., “Decoding the torque of lower limb joints from EEG recordings of pre-gait movements using a machine learning scheme,” *Neurocomputing*, vol. 446, pp. 118–129, 2021.

[3] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[4] Q. Wu, Y. Ge, X. P. Di Ma et al., “Analysis of prognostic risk factors determining poor functional recovery after comprehensive rehabilitation including motor-imagery brain-computer interface training in stroke patients: a prospective study,” *Frontiers in Neurology*, vol. 12, 2021.

[5] C. Li, Y. Xu, L. He, Y. Zhu, S. Kuang, and L. Sun, “Research on fNIRS recognition method of upper limb movement intention,” *Electronics*, vol. 10, no. 11, p. 1239, 2021.

[6] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[7] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, “A CSP\AM-BA-SVM approach for motor imagery BCI system,” *IEEE Access*, vol. 6, pp. 49192–49208, 2018.

[8] S. B. Borgheai, J. McLinden, A. H. Zisk et al., "Enhancing communication for people in late-stage ALS using an fNIRS-based BCI system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp. 1198–1207, 2020.

[9] H. Varsehi and S. M. P. Firoozabadi, "An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality," *Neural Networks*, vol. 133, pp. 193–206, 2021.

[10] Z. Gao, X. Wang, Y. Yang et al., "EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2755–2763, 2019.

[11] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 3839–3852, 2020.

[12] S. Sakhavi and C. Guan, "Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery BCI," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 588–591, Shanghai, China, 2017.

[13] O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğmuş, "Transfer learning in brain-computer interfaces with adversarial variational autoencoders," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 207–210, San Francisco, CA, USA, 2019.

[14] D. Wu, B. Lance, and V. Lawhern, "Transfer learning and active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2801–2807, San Diego, CA, USA, 2014.

[15] D. Wu, "Active semi-supervised transfer learning (ASTL) for offline BCI calibration," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 246–251, Banff, AB, Canada, 2017.

[16] I. Hossain, A. Khosravi, I. Hettiarachchi, and S. Nahavandi, "Calibration time reduction using subjective features selection based transfer learning for multiclass BCI," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 491–498, Miyazaki, Japan, 2018.

[17] E. Jeon, W. Ko, and H.-I. Suk, "Domain adaptation with source selection for motor-imagery based BCI," in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–4, Gangwon, Korea (South), 2019.

[18] Y. Liang and Y. Ma, "Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning," *Biomedical Signal Processing and Control*, vol. 62, article 102101, 2020.

[19] Q. She, B. Hu, Z. Luo, T. Nguyen, and Y. Zhang, "A hierarchical semi-supervised extreme learning machine method for EEG recognition," *Medical & Biological Engineering & Computing*, vol. 57, no. 1, pp. 147–157, 2019.

[20] H. He and D. Wu, "Transfer learning for brain–computer interfaces: a Euclidean space data alignment approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2020.

[21] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[22] D. W. Chen and Y. H. Jin, "An active learning algorithm based on Shannon entropy for constraint-based clustering," *IEEE Access*, vol. 8, pp. 171447–171456, 2020.

[23] M. A. Rahman, F. Khanam, M. Ahmad, and M. S. Uddin, "Multiclass eeg signal classification utilizing rényi min-entropy-based feature selection from wavelet packet transformation," *Brain Informatics*, vol. 7, no. 1, pp. 1–11, 2020.

[24] V. Martínez-Cagigal, E. Santamaría-Vázquez, and R. Hornero, "Asynchronous control of p300-based brain–computer interfaces using sample entropy," *Entropy*, vol. 21, no. 3, p. 230, 2019.

[25] W. Zhang and D. Wu, "Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation," in *2020 international joint conference on neural networks (IJCNN)*, pp. 1–8, Glasgow, UK, 2020.

[26] S. Huang, R. Jin, and Z.-H. Zhou, "Active learning by querying informative and representative examples," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 1936–1949, 2014.

[27] Q. She, K. Chen, Z. Luo, T. Nguyen, T. Potter, and Y. Zhang, "Double-criteria active learning for multiclass brain-computer interfaces," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 3287589, 13 pages, 2020.

[28] W. Zhou and Z. Xu, "Asymmetric fuzzy preference relations based on the generalized sigmoid scale and their application in decision making involving risk appetites," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 3, pp. 741–756, 2015.

[29] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–Graz data set a, Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces)," *Graz University of Technology*, vol. 16, pp. 1–6, 2008.