A dual alignment-based multi-source domain adaptation framework for motor imagery EEG classification

Dong-qin Xu[1] · Ming-ai Li[1,2,3]

Abstract

Domain adaptation, as an important branch of transfer learning, can be applied to cope with data insufficiency and high subject variabilities in motor imagery electroencephalogram (MI-EEG) based brain-computer interfaces. The existing methods generally focus on aligning data and feature distribution; however, aligning each source domain with the informative samples of the target domain and seeking the most appropriate source domains to enhance the classification effect has not been considered. In this paper, we propose a dual alignment-based multi-source domain adaptation framework, denoted DAMSDAF. Based on continuous wavelet transform, all channels of MI-EEG signals are converted respectively and the generated time-frequency spectrum images are stitched to construct multi-source domains and target domain. Then, the informative samples close to the decision boundary are found in the target domain by using entropy, and they are employed to align and reassign each source domain with normalized mutual information. Furthermore, a multi-branch deep network (MBDN) is designed, and the maximum mean discrepancy is embedded in each branch to realign the specific feature distribution. Each branch is separately trained by an aligned source domain, and all the single branch transfer accuracies are arranged in descending order and utilized for weighted prediction of MBDN. Therefore, the most suitable number of source domains with top weights can be automatically determined. Extensive experiments are conducted based on 3 public MI-EEG datasets. DAMSDAF achieves the classification accuracies of 92.56%, 69.45% and 89.57%, and the statistical analysis is performed by the kappa value and t-test. Experimental results show that DAMSDAF significantly improves the transfer effects compared to the present methods, indicating that dual alignment can sufficiently use the different weighted samples and even source domains at different levels as well as realizing optimal selection of multi-source domains.

Keywords Motor imagery · Domain adaptation · Transfer learning · Weighted alignment · Maximum mean discrepancy

1 Introduction

Electroencephalogram (EEG) reflects the spontaneous and rhythmic potential changes generated by neurons in the cerebral cortex, which is widely used in emotion recognition [1–3] and Parkinson's disease detection [4, 5], especially in brain-computer interfaces (BCI) [6–8]. Motor imagery EEG (MI-EEG) based BCI can improve the quality of life for patients with neurological disabilities by translating brain activity directly into command signals through electronic devices. Due to inherent neural activity, concentration level and other factors, brain signals show high inter-subject variability, and aligning the data distribution among subjects has become an important issue in BCI based rehabilitation engineering [9, 10].

The discrepancies among domains can be mitigated by aligning their data distribution according to the decision boundary of the target domain. Ibrahim et al. [11, 12] proposed two data alignment methods based on active learning and instance transfer learning. First, selective informative instance transfer learning (SIITAL) and filter bank common spatial pattern ($SIITAL_{fbcsp}$) were incorporated to investigate the effect of subject-specific feature selection [11]. Next, they proposed an optimal ensemble method to obtain a universal multi-class information instance transfer learning framework based on a combination of previous methods for selecting instance and direct transfer learning [12]. The mentioned methods selected instances of the source domain by applying the decision boundary of the target domain, which increased

✉ Ming-ai Li
limingai@bjut.edu.cn

[1] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

[2] Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China

[3] Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China

the amount of data in the training set and aligned the data distribution between the source and target domains. However, relying entirely on active learning to select instances from the source domain might miss the most similar and effective samples. Wu et al. [13] and Kun et al. [14] used sample reweighting to avoid class imbalance and align the distribution between source and target domains, but it still relied on good initial features. The boosting algorithm can give higher weight to samples that are difficult to learn by adding the weight of each sample [15], while the Kullback–Leibler divergence-based transfer learning algorithm can avoid the problem that poorer samples instead produce higher weights [16]. Compared to sample alignment, researchers have recently focused on feature-level alignment. Yong and Yu [17] proposed a multi-source fusion transfer learning algorithm with the Takagi–Sugeno–Kang fuzzy system (MFTL-TSK) for MI classification, which use Riemannian geometry alignment and balance distribution adaptation to reduce the difference in feature distribution between different subjects. Zhang et al. [18] presented the sub-band target alignment common spatial pattern (SBTACSP) method for cross-subject classification of MI-EEG, which align the source domain trails in each sub-band into the target domain space without changing the distribution of the target domain.

Domain adaptation, as a popular branch of transfer learning, aims to make the distance between the source and target domains with different distributions as close as possible [19–21]. Wang et al. [22] reviewed the recent advances in domain adaptation and domain generalization, and analyzed the generalization problem in depth which improved the development of machine learning. Recently, an increasing number of researchers who engage in BCI, have focused on using domain adaptation to make the most of the available data from source subjects [23–28]. Chai et al. [23] proposed a novel subspace alignment auto-encoder to reduce the difference in data distribution among subjects or sessions, which combined auto-encoder and subspace alignment in a unified framework by using nonlinear transformations and maximum mean discrepancy (MMD). The divergences of marginal and conditional probability distributions between the different domains can be minimized via MMD [24, 25]. Jiang et al. [26] proposed a kernel-based Riemannian manifold domain adaptation framework to align the covariance matrices in the Riemannian manifold, and minimize the conditional distribution distance between the source and target domains based on MMD. Liu et al. [27] proposed a cross-device transfer learning framework based on alignment and pooling for EEG headset domain adaptation, which is accomplished by aligning the spatial pattern and covariance of the source and target domains to realize effective transfer. Peterson et al. [28] proposed backward optimal transport for domain adaptation to boost the performance of an already trained classifier by transforming target samples.

In recent years, deep transfer learning based on domain adaptation has shown the advantage of a high recognition rate [29–31], particularly regarding MI-EEG based BCI rehabilitation systems [32–38]. Jeon et al. [32] devised a multi-path network framework for MI classification, which realized domain adaptation by adapting samples of other subjects and using the gradient reversal layer to update network parameters and improve the network performance. In the same year, they proposed a new domain adaptation method to minimize the distributional discrepancy among subjects by estimating mutual information of subjects in high-level and low-level representation [33]. Moreover, the domain discriminator has gained increasing attention because it can learn deep representations and reduce the discrepancies between domains [34, 35]. Zhao et al. [36] proposed an end-to-end deep domain adaptation method for MI tasks, and the feature distribution shift among domains is matched by the domain discriminator with an adversarial learning strategy. Wei et al. [37] proposed a separate-common-separate network with MMD (SCSN-MMD) for individual subjects and aligned the extracted feature of separate deep feature extractors through the common fully connected layers. Zheng et al. [38] designed a new deep network including an adaptive layer into the full connection layer by minimizing the local MMD and the prediction error to achieve better intra-subject classification. Previous studies have shown the effectiveness of domain adaptation in BCI. However, due to the individual differences and BCI illiteracy, the data distribution of the source domain is different from that of the target domain, resulting in a discrepant auxiliary of the different source domains to the target domain. In addition, it may cause unreasonable use of existing data and insignificant improvement in model performance to extract common domain invariant representations for all source domains or select only one source domain for transfer learning. Therefore, domain adaptation can solve the problem of differences among domains, and it is particularly necessary to use the knowledge of multiple source domains sufficiently and reduce the data distributional discrepancy of multi-source domains and target domain in BCI.

In this paper, a dual alignment-based multi-source domain adaptation framework (DAMSDAF) is proposed by aligning each pair of source and target domains to realize the transfer learning of multiple source domains. First, each source domain is pre-aligned by assigning weights, which are the normalized mutual information (NMI) between the informative sample of the target domain and each sample of the source domain. Second, a multi-branch deep network (MBDN) is designed based on MMD and weighted prediction to align the source and target domains in common feature space. Finally, MBDN with a sequential selection algorithm accomplishes domain-specific data distribution alignment and achieves the optimal transfer effect of multi-source domains.

The main contributions of the work in this paper are summarized as follows:

1. We propose a dual alignment approach to reduce the distribution discrepancies between the source and target domains. To the best of our knowledge, this work is the first to combine weight assignment-based sample-level domain adaptation and MMD-based feature-level domain adaptation for MI-EEG classification.
2. We introduce a multi-branch deep network (MBDN) to align each pair of source and target domains in a specific space, and ensure that each source domain has maximum auxiliary to the target domain via weighted prediction. MBDN is combined with a sequential selection algorithm to select optimal multiple source domains, realizing multi-source transfer learning.
3. We validate that the data distribution is different among different subjects based on the marginal distribution, which means that different subjects can be treated as different domains.
4. Extensive experiments show that the proposed model achieves excellent results in MI-EEG classification, and the performance of DAMSDAF is analyzed based on accuracy, confusion matrices, kappa value, and paired t-test.

The rest of this paper is organized as follows: The related work is introduced in Section 2. In Section 3, the details of the materials and method are described. Section 4 presents the weight of the source domain, the selection of sub-neural network architecture and data visualization. Section 5 describes the experimental evaluation results, comparison of the related work and statistical analysis to validate the effectiveness of our method. Section 6 discusses the results, advantages and limitations. Finally, conclusions are drawn and future work is discussed in Section 7.

2 Related work

This section introduces some basic concepts of transfer learning, entropy and maximum mean discrepancy, and state-of-the-art domain adaptation approaches in MI-EEG based BCIs, which motivated the proposed DAMSDAF.

Considering that a large number of symbols and abbreviations are involved in this paper, we provide a unified explanation in Table 1, and the corresponding descriptions are provided when they first appear in the text, while the abbreviations of the compared methods are described in the corresponding position.

2.1 Transfer learning

Transfer learning is a new machine learning methodology that can solve problems in related but different domains by using existing knowledge [19, 20]. We define transfer learning as follows.

Definition 1 Given a source domain D_S, learning task T_S, target domain D_T, and learning task T_T, transfer learning is dedicated to using the knowledge in D_S and T_S to help improve the learning of the target prediction function $f_T(\cdot)$ in D_T. Here, $D_S \neq D_T$ and/or $T_S \neq T_T$.

In the above definition, where the source domain $D_S = \{\mathcal{X}_S, P_S(X)\}$ and target domain $D_T = \{\mathcal{X}_T, P_T(X)\}$, $\mathcal{X}$ is the feature space, $P(X)$ is the marginal probability distribution and $X = \{x_1, x_2, \cdots, x_n\} \in \mathcal{X}$, learning task $T_S = \{\mathcal{Y}_S, f_S(\cdot)\}$ and $T_T = \{\mathcal{Y}_T, f_T(\cdot)\}$, $\mathcal{Y}$ is the label space. Thus, the condition that the two domains are different implies that either $\mathcal{X}_S \neq \mathcal{X}_T$, $P_S(X) \neq P_T(X)$, $\mathcal{Y}_S \neq \mathcal{Y}_T$ or/and $f_S(\cdot) \neq f_T(\cdot)$. Based on the different situations between the source and target domains and tasks, transfer learning can be categorized into three settings: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning [19]. In this paper, we only consider the condition of $P_S(X) \neq P_T(X)$ in the MI-EEG domain adaptation.

2.2 Sample-level domain adaptation

Domain adaptation is a field associated with machine learning and transfer learning, which has the same feature space between the source and target domains but different distributions, i.e., a subcategory of transfer learning [39, 40]. Table 2 shows the distinction among usual machine learning, transfer learning and domain adaptation.

The main goal of domain adaptation is to minimize the distribution difference between the source and target domains. Sample-level domain adaptation learns a set of model parameters by weighting each sample in the source or selecting the samples with similar distributions, so that the distribution of the source domain approximates the distribution of the target domain [11–16]. Entropy is used as the measurement to select uncertain samples that have more informativeness to learn the decision boundary than other samples [12]. The calculation formula of the entropy is as follows:

$$H_n(i) = -\sum_{c=1}^{n_c} P(y_c|x_i) log_b P(y_c|x_i), i = 1, 2, \cdots, N \tag{1}$$

where N is the total number of samples, n_c is the number of classes, $P(y_c|x_i)$ is the probability of samples x_i being in class y_c, b is the base of the logarithm used and $b = 2$, and the samples were transferred that have entropy equal to or greater than 0.29228. Similarly, the normalized entropy is applied to find informative samples from the source to the target domain [11] using:

$$NH_n = \frac{H_n - min(H_n)}{max(H_n) - min(H_n)} \tag{2}$$

where $min(H_n)$ and $max(H_n)$ are the minimum and maximum of all entropy values, respectively. Therefore, the normalized

Table 1 Notations used in this paper

Notations		Descriptions	Notations	Descriptions
Abbreviations	ASD	Aligned source domain	MBDN	Multi-branch deep network
	BCI	Brain-computer interface	MI	Motor imagery
	CWT	Continuous wavelet transform	MMD	Maximum mean discrepancy
	DAMSDAF	Dual alignment-based multi-source domain adaptation framework	NMI	Normalized mutual information
	DASSDAF	Dual alignment-based single-source domain adaptation framework	RKHS	Reproducing kernel Hillbert space
	EEG	Electroencephalogram	WA	Weighted average
	JS	Jensen-Shannon divergence	WSD	Weighted source domain
Symbols	$C(\cdot)$	Classifier	R_{sl}	Sample size of lth ASD
	D	Domain	S_i	Source subject i
	$\widehat{D}_{\mathcal{H}}(\cdot)$	Distance of MMD	$S_{S_{ka}}$	the kth ASDs in the same dataset
	D_{T_T}	Training set of the target domain	T_S	Target domain
	$F(\cdot), Q(\cdot)$	Feature extractor	W_l	Classifier' weight
	$H(\cdot)$	Entropy	W_{Sample}^l	Sample's weight
	$I(\cdot, \cdot)$	Mutual information	$\mathbf{x}, y$	Samples/ label
	$\mathcal{L}(\cdot, \cdot)$	Loss function	χ, y	Feature/label space
	M	Number of ASD	γ	Adaptation factor
	$P(\cdot)$	Distribution	t	Value of test statistic

entropy value of the sample is in the range of [0.0, 1.0], and the samples having normalized entropy in [0.5,1.0] are transferred.

Sample reweighting is widely applied to achieve distribution adaptation for sample importance and the number of samples of different classes between source and target domains. Kun et al. [14] proposed augmentation-based source-free adaptation (ASFA) to assign different weights to the samples by using the entropy of the rescaled probability with Laplace smoothing. Wu et al. [15] proposed a weighted adaptation regularization algorithm for online and offline, which made distributions of difference classes more consistent between the source and target domain by weighting samples, and the ith sample in the source and target domains as follows:

$$w_{s,i} = \begin{cases} 1, & x_i \in D_{s,1} \\ n_1/(n-n_1), & x_i \in D_{s,2} \end{cases} \tag{3}$$

$$w_{t,i} = \begin{cases} 1, & x_i \in D_{t,1} \\ m_1/(m-m_1), & x_i \in D_{t,2} \end{cases} \tag{4}$$

where $D_{s,c} = \{x_i | x_i \in D_s \wedge y_i = c, i = 1, \cdots, n\}$ and $D_{t,c} = \{x_j | x_j \in D_t \wedge y_j = c, j = n+1, \cdots, n+m_l\}$ are the set of samples in class c of the source domain and target domain, respectively, n_c is the number of elements in $D_{s,c}$, and m_c is the number of elements in $D_{t,c}$.

2.3 Feature-level domain adaptation

Feature-level domain adaptation maps samples from different domains to the same feature space and minimizes distribution discrepancies, which has received extensive attention and developed well [23–26]. MMD [37, 38] is the most frequently used metric distance in domain adaptation, which estimates the discrepancy between two distributions P and Q in a reproducing kernel Hilbert space (RKHS). Formally, MMD is defined by

$$D_{\mathcal{H}}(P,Q) \triangleq \left\| E_X[\phi(x^S)] - E_Y[\phi(x^T)] \right\|_{\mathcal{H}}^2 \tag{5}$$

where $\mathcal{H}$ is the RKHS endowed with a characteristic kernel k, $\phi(\cdot)$ is the feature map that maps the original samples to RKHS, kernel k means $k(x^S, x^T) = \langle \phi(x^S), \phi(x^T) \rangle$, and $\langle \cdot, \cdot \rangle$ is the inner product of vectors.

In deep learning, the computed MMD between the source and target domains in the feature extractors is added to the loss function. Wei et al. [37] designed a separate-common-separate network by separating the feature extractor of the convolutional neural network for subjects, including three fully connected layers to extract common features of all subjects in the feature space. Furthermore, they computed MMD for each of the three layers cross-subject to increase the

Table 2 Distinction among usual machine learning, transfer learning and domain adaptation

Learning paradigms		Conditions	
		$P_S(X)$ and $P_T(X)$	T_S and T_T
Machine learning		Same	Same
Transfer learning	Inductive transfer learning	Same	Different
	Transductive transfer learning	Different	Same
	Unsupervised transfer learning	Different	Different
Domain adaptation	Sample-level domain adaptation	Different	Same
	Feature-level domain adaptation	Different	Same

significance of deeper layers, where the MMD loss is a weighted average MMD of each of the three layers and averaging weights are 1/6, 1/3 and 1/2, respectively. Zheng et al. [38] aggregated the same class of EEG data by minimizing the local MMD between different domains, and the form of the local MMD is as follows:

$$D_{\mathcal{H}}(P,Q) = \frac{1}{C} \sum_{c=1}^{C} \left\| \sum \omega_i^{Sc} \phi(x^S) - \sum \omega_i^{Tc} \phi(x^T) \right\|_{\mathcal{H}}^2 \tag{6}$$

where $\sum \omega_i^{Sc} = 1$, $\sum \omega_i^{Tc} = 1$, and ω^c represents the proportion of experimental data belonging to class c. Then, the local MMD loss obtained by the adaptive layer, which is after the full connection layer of the deep network, was added to the total loss function to train the model. In the specific calculation process of the abovementioned MMD distance, the corresponding **x** was replaced by the activation function value in the full connection. In this work, we demonstrate a novel MBDN, which embeds MMD in each branch to align each pair of source and target domains and improve the performance of the model.

3 Materials and method

This section introduces the proposed DAMSDAF in detail. Figure 1 shows the flowchart of DAMSDAF. First, the raw MI-EEG signal is transformed to acquire time-frequency spectrum images based on a continuous wavelet transform. Second, each source domain is weighted by assigning the NMI, which is obtained with the informative samples of the target domain, and the weighted source domain (WSD) is transferred to the training set of the target domain (D_{T_T}) to form the aligned source domain (ASD). Third, domain-specific distribution alignment is achieved via MMD, and the sub-classifiers are weighted with the single branch transfer accuracy of the validation set to obtain the prediction of the test set of the target domain. Finally, the sequential selection algorithm with MBDN is used to realize the multi-sources' optimization and enhance the transfer effect of the multi-source domains to the target domain.

3.1 Datasets source and preprocessing

We used three public MI datasets to evaluate the performance of DAMSDAF. The details of these datasets are given below.

Dataset 1 (Dataset 2b from BCI Competition IV [41]): This dataset was recorded from 9 healthy subjects. Each subject performed MI of the left and right hands, which was made up of the 3 EEG channels (C3, Cz and C4) and 3 EOG channels, and the sampling rate was 250 Hz. For each subject, 5 sessions were provided and each session consisted of 120, 140 or 160 9 s trials. The first three sessions were set as the training

set, whereas the remaining two sessions were set as the testing set. The 3 s–7 s signals of each trial from the 3-channel EEG data were chosen for subsequent experimental study.

Dataset 2 (Dataset 2a from BCI Competition IV [42]): This dataset consisted of 9 healthy subjects, each subject performed four different mental tasks of MI of feet, left hand, right hand and tongue. 22 channels (Ag/AgCl) were used for EEG signal recording and 3 channels were used for EOG recording. Both the EEG and EOG channels were sampled at 250 Hz. Each subject completed training and an evaluation session. In each session, a subject performed 72 trials per class, which was 288 in total. We only used the 3-channel EEG data (C3, Cz and C4), and the 3 s–6 s signal of each trial was extracted as the experimental data in this paper.

Dataset 3 (Dataset III from BCI competition II [43]): This dataset was recorded from a healthy female, 25 y subject, and it was made up of left- and right-hand MI-EEG. 3 channels (C3, Cz and C4) with a sampling frequency of 128 Hz were applied to record EEG signals. There are 280 trials of 9 s length in the dataset, in which the training and testing sets were both 140 times. For each EEG trial, the MI tasks with a duration between 3 s and 9 s were used for the experiment.

In the specific experimental process, the results of datasets 1 and 2 are calculated using the leave one subject out (LOSO) validation method, i.e., nine subjects in datasets 1 and 2 are successively taken as the target subject, and the remaining eight subjects are taken as multiple source subjects. In addition, the experiment is also studied on cross-dataset transfer learning, where nine subjects of dataset 1 are used as multiple source subjects and one subject of dataset 3 was taken as the target subject. After preprocessing, the trials of the target subject are split into training, validation and test sets. For the target subject, the training set is the training set marked in the datasets, the validation set contains the first half trials of the testing set marked, and the last half trials of the testing set marked from the test set. In addition, the training set marked in the datasets is used as the trials of the source subject.

3.2 Time-frequency spectrum image generation based on CWT

The variation in energy is often caused by the potential activity of the contralateral cortex and ipsilateral cortex during MI, and this phenomenon is not clearly reflected in the time domain. To describe the features in a better form, CWT is used to transform the time series signals into two-dimensional images. First, CWT is applied to MI-EEG signals of each channel, and the bandwidth parameter and center frequency of the complex Morlet wavelet are kept at 3, in which the frequency spectrum corresponding to the 0–32 Hz frequency band is extracted as the feature representation of MI-EEG in the time and frequency domains. Then, the images of the C3, Cz and C4 channels

Fig. 1 Flowchart of DAMSDAF
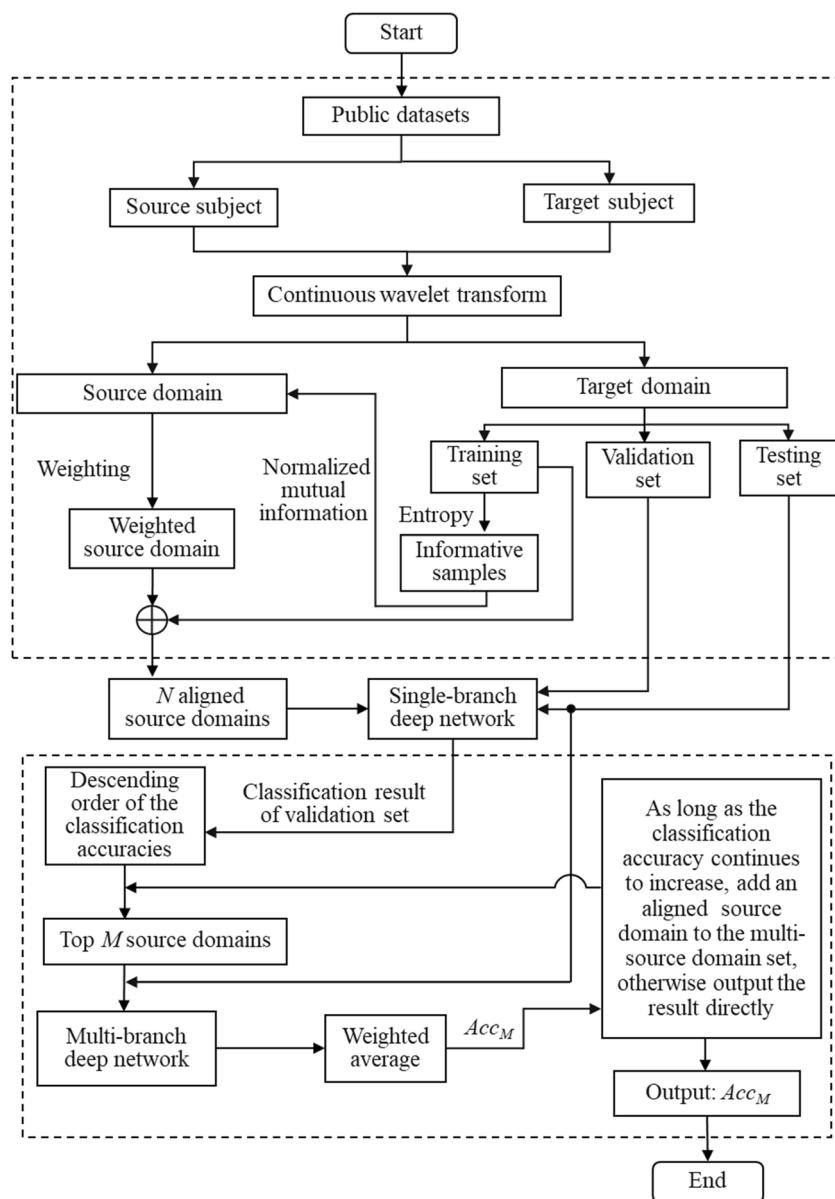
are vertically stacked into a new image, which is resized to 224×224 to form time-frequency spectrum images and construct multi-source domains and the target domain. The detailed conversion process of the raw MI-EEG signal to time-frequency spectrum images is shown in Fig. 2.

### 3.3 Align the source and target domains based on weight assignment

The data distribution of the source domain directly affects the transfer effect, and the distance between the source and target domains can be narrowed by weighting the source domain. Figure 3 shows the weighted alignment process for the source and target domains. First, the decision boundary of D_{T_T} is obtained by using pre-trained ResNet50, and the informative samples are selected from D_{T_T} based on entropy. Then, NMI

is calculated between informative samples and time-frequency spectrum images of the source domain and is assigned to the source domain as weights, and ASD is formed of WSD and D_{T_T}.

According to the active learning method, the samples close to the decision boundary have the most uncertainty, while the uncertain samples have more information than others to learn the decision boundary, which can accelerate the learning process of the model. Therefore, the informative samples near the decision boundary of D_{T_T} are selected according to the normalized entropy by Formula (2), which can quantify the information carried by samples, and the time-frequency spectrum images with high entropy values are selected as informative samples.

Then, the weight assignment for time-frequency spectrum images of the source domain is accomplished based on NMI.

Fig. 2 The conversion process of raw EEG signals to time-frequency spectrum images based on CWT. The raw EEG signals of the C3, Cz and C4 channels are transformed into two-dimensional images by CWT, and the 0–32 Hz frequency band is extracted as feature images, which are vertically stacked and resized to 224 × 224 to form time-frequency spectrum images. S_i denotes the source subject, $i = 1, 2, \cdots, N$, T denotes the target subject
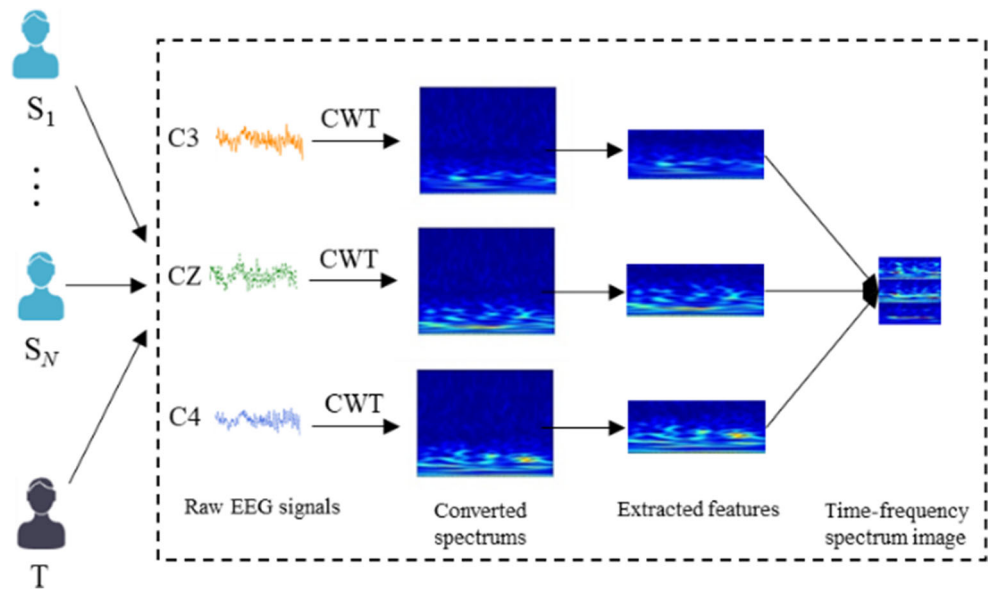
In probability and information theory, the mutual information of two random variables is a measure of the mutual dependence between the two variables.

Definition 2 Let (X, Y) be a pair of random variables with values in the space $\mathbf{X} \times \mathbf{Y}$. If their joint distribution is $\rho(x, y)$ and the marginal distributions are $\rho(x)$ and $\rho(y)$, the mutual information of two jointly discrete random variables X and Y is calculated as a double sum:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} \rho(x,y) log \frac{\rho(x,y)}{\rho(x)\rho(y)} \tag{7}$$

The normalized variant of the mutual information [44] is given by

$$NMI(X;Y) = 2 \frac{I(X;Y)}{H(X) + H(Y)} \tag{8}$$

where $H(X)$ and $H(Y)$ are the marginal entropies of random variables X and Y, respectively. In this paper, Formula 4 is used to calculate the NMI value between time-frequency spectrum images of the source domain and informative samples, which are assigned for the time-frequency spectrum images of the source domain as weights. Then, WSD is transferred to
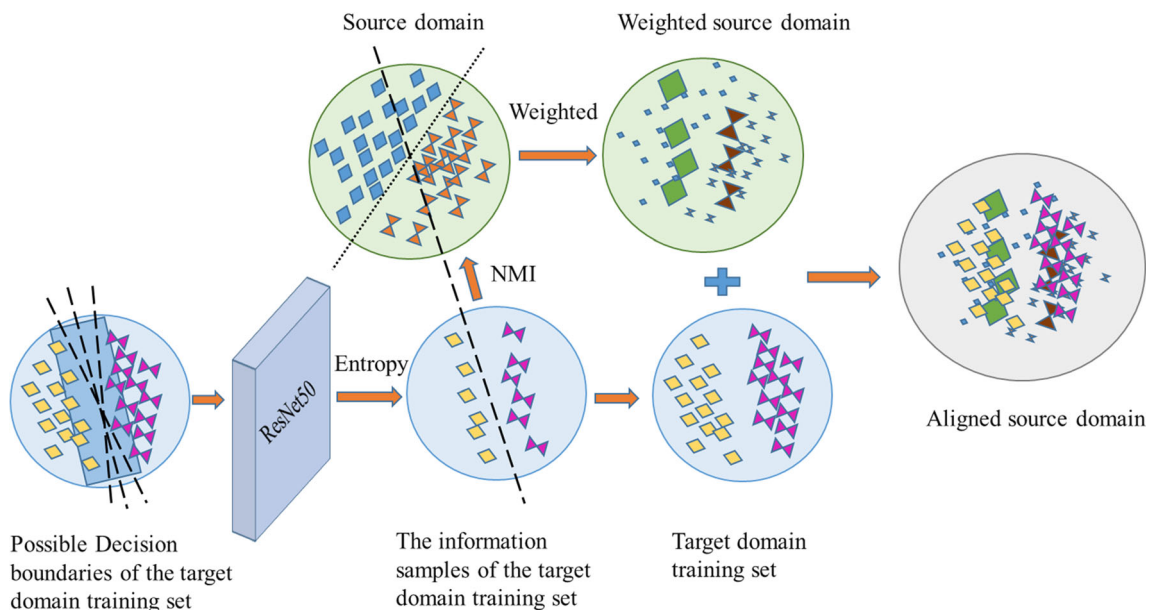
Fig. 3 Schematic diagram of the weighted alignment process for the source and target domains

D_{T_T} and constitutes ASD, which not only increases the number of training samples but also aligns the data distribution in different domains and lays the foundation for the next work.

3.4 Domain-specific distribution alignment and weighted prediction

The specific features of each source domain are different, and the adaptability among domains can be increased by minimizing the distance between each pair of ASDs and the target domain. First, multiple ASDs ($S_{S_l}, l = 1, 2, \cdots, M$) and target domain ($T_S$) are input into the shared network to extract the common potential representations, which are respectively fed into the sub-neural network to extract domain-specific features. Then, the domain-specific features extracted from the source and target domains are aligned via MMD. Finally, the test set of the target domain is classified by weighted prediction based on the MBDN. The detailed architecture of the MBDN is shown in Fig. 4.

The MBDN consists of three parts: a shared network, sub-neural networks and sub-classifiers. For details, a pre-trained ResNet50 based on ImageNet is used as the shared network $F(\cdot)$ to extract the common potential representations of all domains; then, the common features are fed into the unshared lth sub-neural network $Q_l(\cdot)$; thus, each pair of ASD and target domain are mapped into a specific feature space to obtain domain-specific features $Q_l(F(X_{sl})$)and $Q_l(F(X_t)$), where X_{sl} is the lth source domain and X_t is the target domain. The structure of sub-neural network is shown in the blue dotted box in Fig. 5. The sub-classifier $C_l(\cdot)$ is a softmax classifier, which is shown in the last light purple block in Fig. 5. The number of neurons is equal to that of types of MI tasks in the target domain. In addition, a batch normalization layer is applied after each convolutional layer to speed up the training process, and a dropout layer is applied in each sub-network to prevent overfitting (dropout = 0.5). In this work, the shared network can also be obtained by fine-tuning the models of other ResNet. The specific training process of the MBDN is to fine-tune all convolutional and pooling layers of the shared

network to learn the common representations for all domains, while the sub-neural networks are trained from scratch with each ASD to learn domain-specific representations, which can improve the generalization ability of the proposed model.

To achieve the data distribution alignment of each pair of source and target domains in a specific space, MMD is applied as the estimation of the inter-domain discrepancy in this paper. According to Formula (5), MMD defines the distance between ASD and the target domain.

$$\widehat{D}_{\mathcal{H}}(P,Q) = \left\| \frac{1}{R_{sl}} \sum_{x_j \in D_s} \phi\left(x_j^l\right) - \frac{1}{R_t} \sum_{x_i \in D_t} \phi\left(x_i^t\right) \right\|_{\mathcal{H}}^2$$
$$= \left[\frac{1}{R_{sl}^2} \sum_{j,i=1}^{R_{sl}} k\left(x_j^l, x_j^l\right) - \frac{2}{R_{sl}R_t} \sum_{j,i=1}^{R_{sl},R_t} k\left(x_j^l, x_i^t\right) + \frac{1}{R_t^2} \sum_{j,i=1}^{R_t} k\left(x_i^t, x_i^t\right) \right]^{\frac{1}{2}} \quad (9)$$

where $x_j^l \in X_{Sl}$ and $x_i^t \in X_t$, $\phi(\cdot)$ is feature mapping, i.e., mapping the original samples to RKHS, k is the characteristic kernel, and R_t and R_{sl} are the sample sizes of the target domain and lth source domain, respectively. For domain adaptation of multiple source domains, the MMD loss is reformulated as:

$$\mathcal{L}_{MMD} = \frac{1}{M} \sum_{l=1}^{M} D_{\mathcal{H}}\left(Q_l\left(F(X_{sl}), Q_l(F(X_t)\right)\right) \quad (10)$$

For the lth source domain, the domain-specific invariant features, which are obtained through the lth sub-neural network, are received by $C_l(\cdot)$. For each classifier, the cross entropy is used as the classification loss, and the formula is:
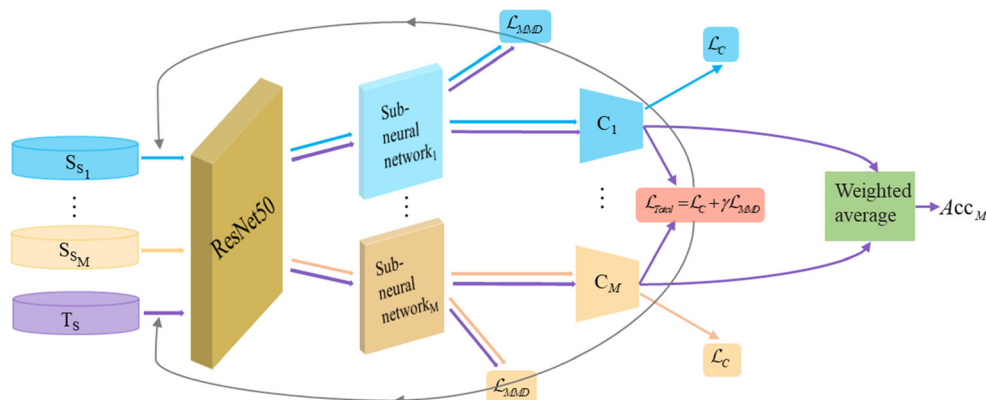
$$\mathcal{L}_C = \sum_{l=1}^{M} W_{Sample}^l J\left(C_l\left(Q_l\left(F\left(x_j^l\right)\right)\right), y_j^l\right) \quad (11)$$

where W_{Sample}^l and $\mathbf{y}_j^l \in Y_{Sl}$ are the sample weight and label of the lth ASD, respectively.

$$W_{Sample}^l = \begin{cases} NMI\ value, & x_g^l \in WSD_l, g \in 1, 2, \cdots, n \\ 1, & x_r^T \in D_{T_T}, r \in n+1, n+2, \cdots, n+m \end{cases} \quad (12)$$
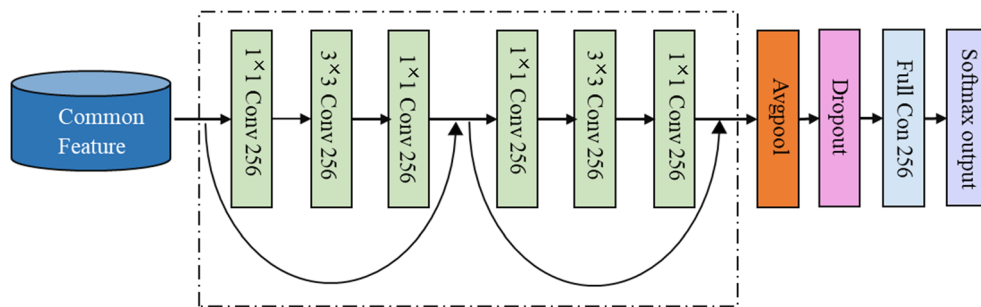
where n and m are the number of samples in WSD_l and D_{T_T}, respectively, and $n + m = R_{sl}$. Then, the loss of DAMSDAF

Fig. 4 Architecture of multi-branch deep network

Fig. 5 Architecture of sub-neural network and sub-classifier

consists of MMD loss and classification loss. The MBDN can accurately classify the ASD data by minimizing classification loss and aligning domain-specific features by minimizing MMD loss. The total loss is formulated as follows:

$$\mathcal{L}_{Total} = \mathcal{L}_C + \gamma \mathcal{L}_{MMD} \qquad (13)$$

To suppress noisy activations at the early stages of training, instead of fixing the adaptation factor γ, as the strategy proposed in [45], we gradually change it from 0 to 1 by a progressive schedule:

$$\gamma = \frac{2}{1 + \exp(-\theta \cdot \zeta)} - 1 \qquad (14)$$

where $\theta = 10$ is fixed throughout the experiments, $\zeta = \delta/U$ is the training progress linearly changing from 0 to 1, $\delta = 1, 2, \cdots, U$, and U is the number of training iterations. This progressive strategy stabilizes parameter sensitivity for DAMSDAF.

Because each sub-classifier has a different prediction for the test set of the target domain, directly averaging the output

of all sub-classifiers does not always yield the expected results. Hence, the optimal decision probability of the target domain is obtained by weighting multiple sub-classifiers:

$$P_T^c(x_t) = \sum_{l=1}^{M} W_l * P_l^c(x_t) \qquad (15)$$

where P_l^c is the probability generated by the lth sub-classifier for class c and W_l is the weight for the corresponding sub-classifier. The weight W_l is given as follows:

$$W_l = \frac{Acc_{Sl}}{\sum_{l=1}^{M} Acc_{Sl}} \qquad (16)$$

where Acc_{Sl} is the classification accuracy of the validation set of the lth single-source to target transfers, which is obtained by using the dual alignment-based single-source domain adaptation framework (DASSDAF), and the output probability is optimized by assigning weights for each sub-classifier. The procedure of domain-specific distribution alignment and weighted prediction is summarized in Algorithm 1.

Algorithm 1 Domain-specific distribution alignment and weighted prediction

---

Input: ASDs, target domain validation and test sets

Output: The accuracies of validation and test sets of target domain

Initialization:

The number of training iterations U

The number of source domains M, $l = 1, 2, \cdots, M$

for δ in 1: U //   Training multi-branch deep network

    1: Feed $ASDs$ and D_{T_E} to shared network to get the common representations $F(x_j^l)$ and $F(x_t)$

    2: Feed $F(x_j^l)$ to SNN to get domain-specific features of source domain $Q_l(F(x_j^l))$

    3: Feed $F(x_t)$ to all SNNs to get $Q_1(F(x_t)), \cdots, Q_M(F(x_t))$

    4: Use $Q_l(F(x_j^l))$ and $Q_l(F(x_t))$ to calculate MMD loss using Formula (10)

    5: Feed $Q_l(F(x_j^l))$ to sub-classifier to get $C_l(Q_l(F(x_j^l)))$,

       and the classification loss is computed using Formula (11)

    6: Update the shared network $F(\cdot)$, multiple SNNs Q_l

       and multiple sub-classifiers C_l by minimizing the total loss in Formula (13)

    7: Weighted the multiple sub-classifiers C_l with W_l //   Weighted prediction

    8: Output the accuracies of test set Acc_M

end

3.5 Sequential selection of multiple source domains

Multiple source domains with preferable transfer effects are selected for simultaneous transmission which can ensure all transferred source domains are positively transferred to the target domain. The sequential selection algorithm for multi-source domains transfer learning is shown in Algorithm 2, which is an improvement of the sequential forward floating-point search algorithm, where Acc denotes the classification accuracy. In each

loop, an ASD is successively added to the current subset as long as the resulting subset is superior to the previously evaluated, and the process continues until the classification accuracy of the target domain no longer increases. To reduce the computational cost, the classification accuracies of all single-source to target transfers are obtained by applying DASSDAF and arranged in descending order (ASD_1, $\cdots$, ASD_N). Then, the corresponding ASD is added in turn to form the optimal multi-source domains.

Algorithm 2 Sequential selection of multiple source domains

Input: A subset of whole ASDs, target domain validation and test sets

Output: The optimal multiple source domains and the *Acc* of test set of target domain

Initialization:

    The accuracy of target domain $Acc = 0$

                               N

                  M

    The accuracy of target domain Acc_M

    Stop when the accuracy does not increase $Acc \geq Acc_M$

while $M \leq N$

  1: Subset of whole ASDs $S_{S_{1,\cdots,M}}^A = \{ASD_1, \cdots, ASD_M\}$

  2: Feed $S_{S_{1,\cdots,M}}^A$ and target domain to multi-branch deep network,

    obtained the classification accuracy of target domain Acc_M

  3: Compare the current Acc_M and the previous Acc

    if $Acc \geq Acc_M$

      M+=1   //   Added a ASD and start a new loop

      if $Acc \geq Acc_M$ and $Acc_M \geq Acc_{M+1}$ output Acc and optimal multiple source domains $S_{S_{1,\cdots,M}}^A$

        break   //   Break out of the while loop

    else

      $Acc = Acc_M$

      M+=1   //   Added a ASD and start a new loop

  end

---

4 Experiment

Three public MI datasets were used to evaluate the proposed DAMSDAF in our experiments. All experiments were carried out with the same software (Spyder, Windows 10) and hardware (a Hewlett-Packard computer, equipped with an Intel(R) Core (TM) i7–9700 CPU @ 3.00 GHz, an NVIDIA GeForce RTX 2070 GPU).

## 4.1 Correlation between domains

To verify that the two domains are different in cross-dataset transfer learning, Jensen–Shannon divergence (JS) is used to measure the difference between the data

distribution of the source domain and target domain based on the marginal distributions $P_S(X)$ and $P_T(X)$, which is symmetric and suitable for measuring differences in data distribution between domains. It is defined by

$$JS\left(P_S(X)\|P_T(X)\right) = \frac{1}{2}D_{KL}\left(P_S(X)\|M\right) + \frac{1}{2}D_{KL}\left(P_T(X)\|M\right) \quad (17)$$

where $D_{KL}(P_S(X)\|M)/D_{KL}(P_T(X)\|M)$ is the Kullback–Leibler divergence and $M = (P_S(X) + P_T(X))/2$. The value of JS is in the range of [0.0, 1.0], where 0 indicates that the distribution of data between domains is perfectly identical. Figure 6 shows the JS

Fig. 6 JS of the target and source domains on cross-dataset transfer learning. *Si* is one of nine source domains from dataset 1, and D_T is the target domain from dataset 3
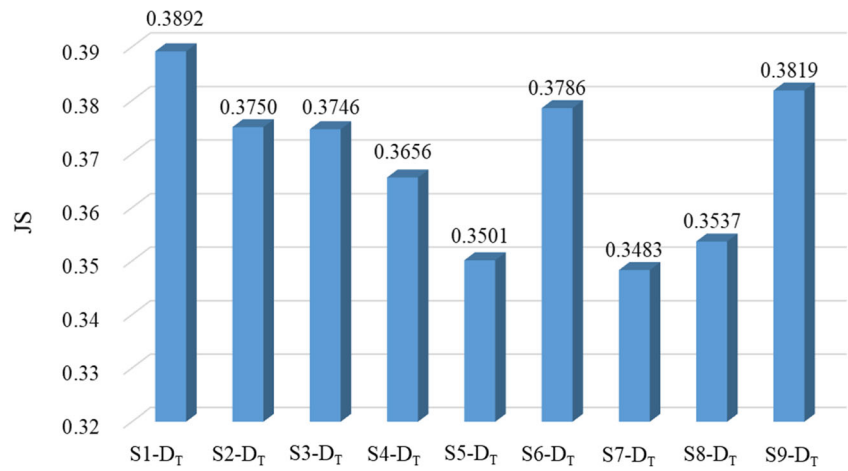

Fig. 7 Weights of different source domains. The blue bars represent the NMI values of the time-frequency spectrum images generated by trials; they are discrepant with each other as well as between different source domains (**a**) Subject 1 (**b**) Subject 2

(a) Subject 1

(b) Subject 2
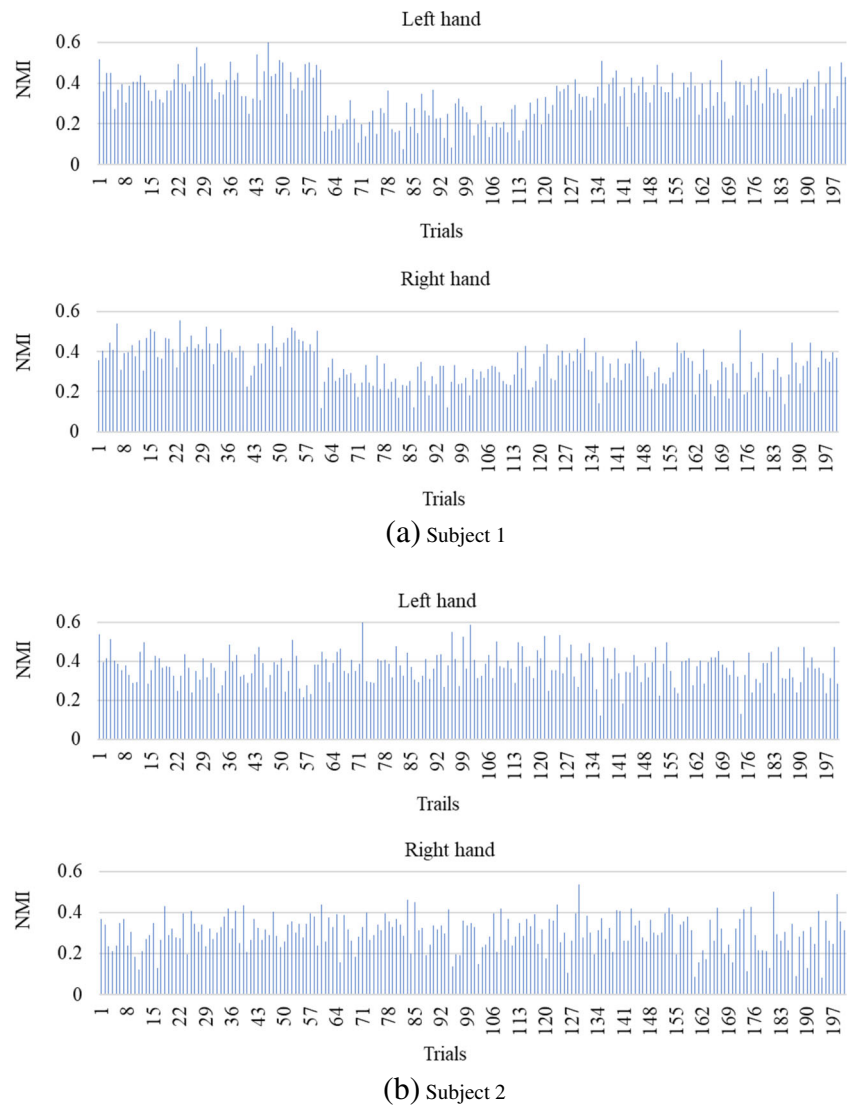
between different source and target domains on cross-dataset transfer learning, where Si is the ith source domain, i = 1, 2, $\cdots$, 9, and D_T is the target domain. The results show that JS between different source and target domains is larger than 0.3, demonstrating that the distribution of data between domains is actually different.

4.2 Weight assignment of the source domain

To maximize the auxiliary of the source domain to the target domain, the source domain is weighted and aligned based on the NMI. First, the informative samples of D_{T_T} are selected according to Formula (2), and then the NMI between the time-frequency spectrum images of the source domain and informative samples is calculated by Formula (8). Figure 7 shows the obtained NMI of Subjects 1 and 2 when Subject 9 is the target subject in dataset 1, where Fig. 7a and b are the NMI of Subjects 1 and 2, respectively. For example, in Fig. 7a, there is a significant difference among NMI of different time-frequency spectrum images. The 46th sample has the maximum NMI of 0.5979, indicating that it is the most relevant to informative samples and conducive to decision-making of the target subject. The 82nd sample has the minimum NMI of 0.0769, which has a low relevance with informative samples and provides little help to the decision-making of the target subject. In addition, the weights between different source domains are also different. As seen from the right hand, the 23rd sample has the maximum NMI of 0.5550 in Subject 1, while the 129th sample has the maximum NMI of 0.5343 in Subject 2, and the average NMI of Subjects 1 and 2 are 0.3350 and 0.3023, respectively, which shows that different source domains have different correlations to the target domain. The other two datasets are processed in the same way. Therefore, each time-frequency spectrum image of the source domain is weighted and transferred to D_{T_T} to form an ASD.
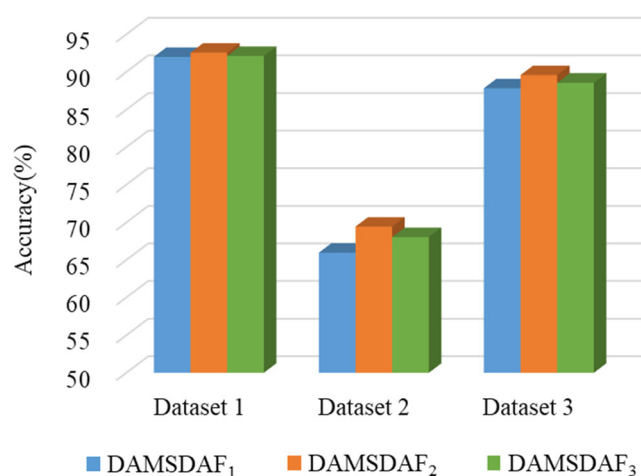
4.3 Determination of sub-neural network architecture

In this paper, MBDN is a modified network based on ResNet50. Due to the great difference between time-frequency spectrum images and ImageNet, pre-trained ResNet 50 with ImageNet has difficulty accurately extracting MI-EEG features. Although a sub-neural network can be trained from scratch, it is not necessarily the most suitable model for MI-EEG classification. Therefore, the same structure (conv (1×1), conv (3×3), conv (1×1)) is called the bottleneck, and the structure of sub-neural network is optimized by changing the number of bottlenecks. The batch size is set as 24 and the number of training iterations U is set to 10,000. Since the sub-neural networks and sub-classifiers are trained from scratch, the learning rate is set to 10 times that of the other fine-tuned layers. In addition, stochastic gradient descent with 0.9 momenta is used as the optimizer, and the learning rate annealing strategy is as follows [45]:

$$\vartheta_p = \frac{\vartheta_0}{(1 + \eta \times \zeta)^\alpha} \tag{18}$$

where ϑ_0 = 0.01, η = 10 and α = 0.75, which are optimized to promote convergence and low error on the source domain.

Figure 8 shows the average accuracies obtained by the MBDN with different structures on 3 datasets, where DAMSDAF$_i$, i = 1, 2, 3, represents the number of bottlenecks in the sub-neural network of the MBDN. DAMSDAF$_1$ achieved the lowest classification accuracy on the 3 datasets, and the gap between DAMSDAF$_2$ and DAMSDAF$_3$ is very marginal on dataset 1, but there is a significant difference on datasets 2 and 3, while the results of 3 datasets all achieved the highest when the sub-neural network structure is DAMSDAF$_2$, indicating that the structure of DAMSDAF$_2$ is optimal, and the detailed architecture of sub-neural network is shown in Fig. 5.

Fig. 8 Classification accuracies of DAMSDAF with different structures on 3 datasets
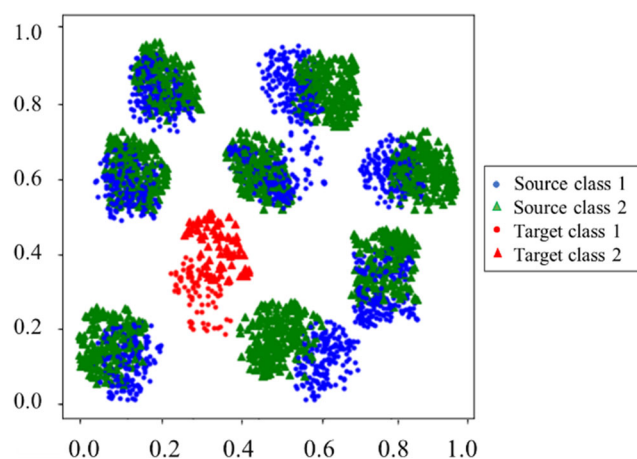
Fig. 9 *t*-SNE visualization of the common features when the first eight subjects and last one subject were used as the source domains and target domain respectively in dataset 1.

Fig. 10 *t*-SNE visualization of the data distributions before and after dual alignment on transferring the optimal source domain (S_4) to the target domain (S_9) in dataset 1 (**a**) Before dual alignment (**b**) After dual alignment.
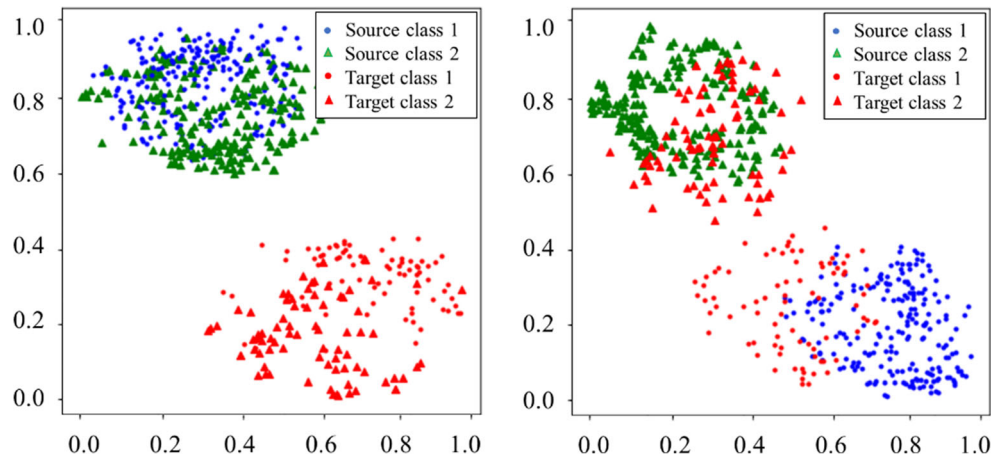
4.4 Visualization

To visually show how the data distribution between the source and target subject was aligned by DAMSDAF, *t*-stochastic neighbor embedding (*t*-SNE) was used for data visualization. Fig. 9 shows the common features that were extracted from nine subjects (S_i, $i = 1, 2, \cdots, 9$) in dataset 1 by using the shared network and normalized, in which S_9 was set as the target domain, and the part labeled data of S_9 and weighted S_n, $n = 1, 2, \cdots, 8$ were set as the multiple ASDs. The blue dots represent class 1 and the green triangles represent class 2 in ASDs, red dots represent class 1, and red triangles represent class 2 in the target domain. It can be seen that the distributions of the target domain and ASDs are quite different.

Next, the domain-specific features, which were extracted via MBDN, are visualized in Fig. 10. In each subplot, the blue dots represent class 1 and the green triangles represent class 2 in the optimal source domain (S_4), while the red dots represent class 1 and the red triangles represent

class 2 in the target domain (S_9), where all data have been normalized. For the data before dual alignment, dots and triangles are indistinguishable, and the source and target domains are very dispersive. After dual alignment, the source and target domains overlap, and the data distribution is clearly visible. By comparing Figs. 9 and 10, it can be seen that samples from the same class in each pair of source and target domains are close, which verifies the effectiveness of DAMSDAF in aligning the data distribution and benefits for subsequent classification.

5 Results and analysis

In this section, the classification performance and robustness of DAMSDAF are verified by presenting the results obtained on datasets 1, 2 and 3. Table 3 summarizes the classification accuracies and kappa values of single-source and multi-source to target transfers on three datasets, where Avg denotes the

Table 3 Classification accuracies (%) and kappa values (accuracy/kappa value) of single-source and multi-source to target transfers on three datasets

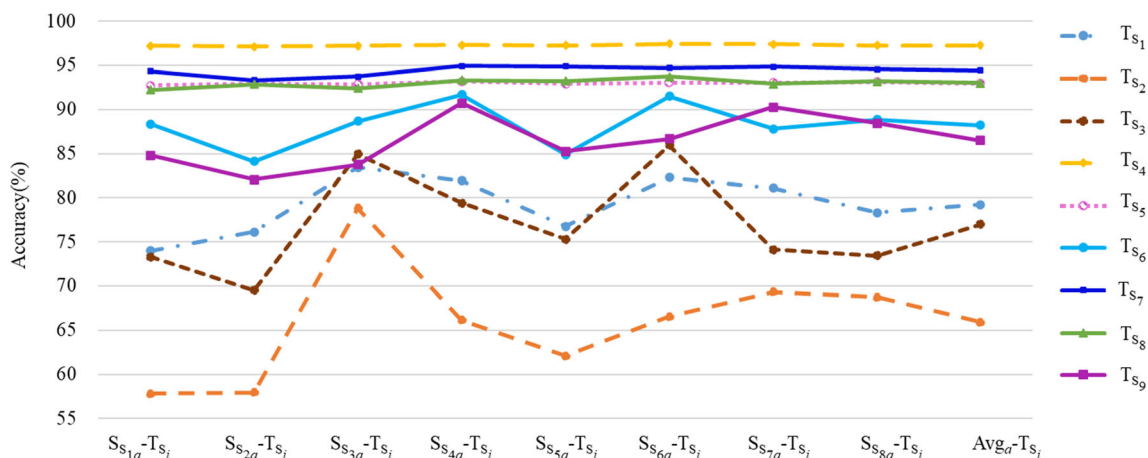| Target subject	Single-source to target transfers			Multi-source to target transfers		
	Inter-subject		Cross-dataset	Inter-subject		Cross-dataset
	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3
T_{S_1}	79.25/0.5850	72.84/0.6379	79.80/0.5960	90.25/0.8050	75.35/0.6713	89.57/0.7914
T_{S_2}	65.92/0.3185	51.40/0.3520	–	87.56/0.7512	54.51/0.3935	–
T_{S_3}	77.00/0.5400	81.63/0.7551	–	88.94/0.7788	83.40/0.7787	–
T_{S_4}	97.27/0.9454	42.26/0.2302	–	97.50/0.9500	45.50/0.2733	–
T_{S_5}	92.99/0.8597	48.10/0.3080	–	93.25/0.8650	52.43/0.3657	–
T_{S_6}	88.24/0.7649	56.53/0.4204	–	93.06/0.8612	58.33/0.4444	–
T_{S_7}	94.41/0.8882	79.57/0.7277	–	95.94/0.9188	83.06/0.7741	–
T_{S_8}	93.00/0.8600	83.74/0.7832	–	94.82/0.8964	86.60/0.8213	–
T_{S_9}	86.50/0.7300	83.28/0.7771	–	91.75/0.8350	85.83/0.8111	–
Avg	86.07/0.7213	66.60/0.5546	79.80/0.5960	92.56/0.8513	69.45/0.5926	89.57/0.7914

Fig. 11 Classification accuracies of different single-source to target transfers on dataset 1

average classification accuracy. In addition, the results are compared with related transfer learning methods to verify the superiority of DAMSDAF.

## 5.1 Inter-subject transfer learning

5.1.1 Dataset 1: BCI IV 2b

To validate that the knowledge transfer effectiveness of multiple source domains is better than that of a single source domain, the classification accuracy is taken as the average of 10 times to negate the random starting sample effect. Figure 11 shows the accuracies of different single-source to target transfers and the average accuracy of each target domain on dataset 1. $S_{S_{ka}}, k = 1, 2, \cdots, 8$ represents the reorder of 8 ASDs after removing the target subject from the dataset, Avg_a represents the average accuracy of the target domain, $T_{S_i}, i = 1, 2, \cdots, 9$ represents 9 target subjects and $S_{S_{ka}}$-T_{S_i} represents the knowledge transfer from the kth ASD to the

ith target domain. The auxiliary of each source domain to the target domain is different, except for T_{S_4}, and it is obviously not optimal to average the transferred results of all single-source to target transfers as the last classification result of the multi-source domains transfer learning. Therefore, selecting the most relevant multiple source domains is crucial to achieving the best transfer effect.

Figure 12 shows the classification accuracies of each target subject through DAMSDAF on dataset 1, i.e., the results of different multi-source to target transfers which are obtained by Algorithms 1 and 2. As seen in Fig. 12, the accuracies of the target subjects show a trend of increasing first and then decreasing or remaining unchanged as the number of source domains increases, reflecting that it is necessary to select the most relevant source domains for transfer learning. In addition, by comparing Figs. 11 and 12, it is found that the transfer effectiveness of multi-source domains is better than that of the single-source domain, i.e., the classification results of multi-source to target transfers are higher than the average of all

Fig. 12 Variation in classification accuracy with an increasing number of source domains on dataset 1
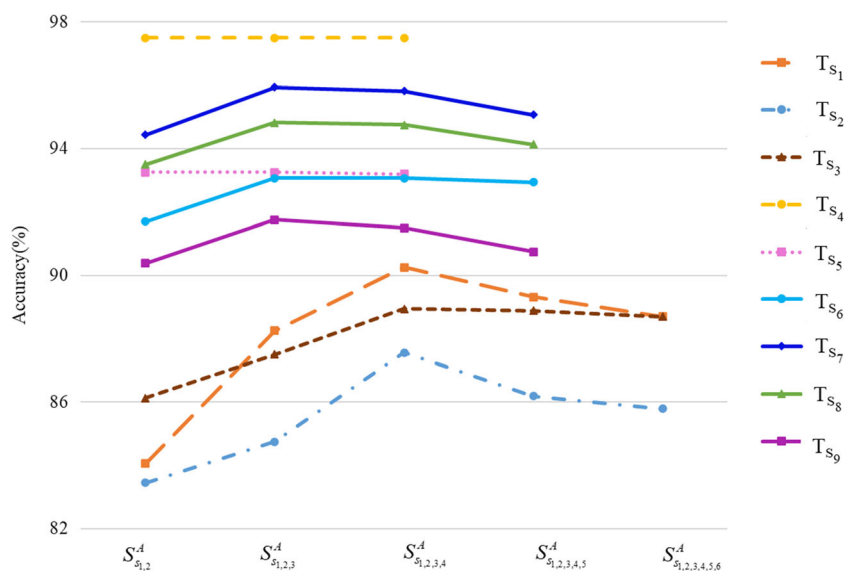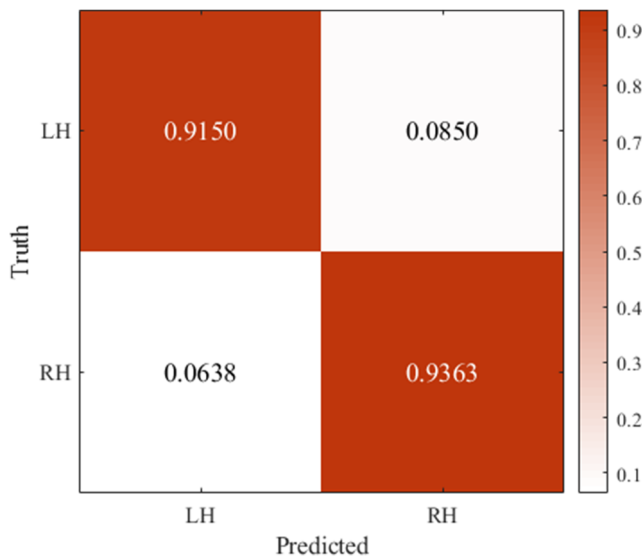
Fig. 13 Confusion matrix obtained with the average of the optimal multi-source to target transfers on dataset 1

single-source to target transfers and the best single-source to target transfers, indicating that the DAMSDAF is effective. Figure 13 shows the confusion matrix of the DAMSDAF method on optimal multi-source to target transfers. LH represents imagining the left-hand task, and RH represents imagining the right-hand task. The classification accuracies of LH and RH are 91.50% and 93.63%, respectively, achieving nearly consistent results, and the accuracies of the two MI tasks are higher than 90.00%, indicating that the proposed method can keep balance and effectively recognize MI tasks of LH and RH on inter-subject transfer learning.

Table 4 shows the comparison of classification accuracies among DAMSDAF, DASSDAF and various relevant transfer learning methods on dataset 1. Here the results of DASSDAF and DAMSDAF are the average of all single-source to target

transfers and optimal multi-source to target transfers classification accuracy, respectively. DAMSDAF is superior to DASSDAF, SIITAL$_{fbcsp}$, SBTACSP and MF-BLDA, and the gaps are 6.50%, 16.77%, 25.71% and 7.86%, respectively. SIITAL$_{fbcsp}$ selected information instances with higher normalized entropy from the source domain by using the trained classifier based on labeled instances of target domain, which may lead to the imbalance of selected samples for all classes or miss the most effective samples. SBTACSP used target alignment to align each frequency sub-band between the source and target domains, and mutual information to select representative CSP features from the sub-bands. MF-BLDA used a boosting algorithm to update the weight of samples in the course of LDA training, which can make the learner focus on the samples with incorrect classification to achieve domain adaptation. DAMSDAF and DASSDAF measured the similarity between the time-frequency spectrum images of the source domain and informative samples by calculating NMI, which was set as the weight coefficient for the time-frequency spectrum images of the source domain to align the data distribution of the source domain. In addition, the kappa value is a statistical measure, that can verify the consistency of DAMSDAF, which is calculated on the classification results based on LOSO validation in this paper. The results are also shown in Table 4, which comprises the kappa values of various transfer learning methods on dataset 1. The 9 subjects had an average kappa value of 0.8513, and there was no significant difference in all subjects, indicating that the DAMSDAF has a strong generalization ability. To summarize, the kappa values of the 9 target subjects and average are all higher than those of the compared transfer learning methods on dataset 1, which strongly proves that DAMSDAF has good robustness and consistency. Hence, the results in Table 4 illustrate that DAMSDAF has better applicability to select and align the source domain based on

Table 4 Comparison of the accuracies (%) and kappa values (accuracy/kappa value) of DAMSDAF with different transfer learning methods on dataset 1

| Target subject	Transfer learning methods				
	SBTACSP [18]	SIITAL$_{fbcsp}$ [11]	MF-BLDA [15]	DASSDAF	DAMSDAF
T_{S_1}	72.50/0.4500	70.28/0.4056	79.40/0.5880	79.25/0.5850	90.25/0.8050
T_{S_2}	66.33/0.3266	61.71/0.2342	80.30/0.6060	65.92/0.3185	87.56/0.7512
T_{S_3}	60.00/0.2000	57.00/0.1400	70.30/0.4060	77.00/0.5400	88.94/0.7788
T_{S_4}	85.83/0.7166	94.40/0.8880	95.00/0.9000	97.27/0.9454	97.50/0.9500
T_{S_5}	58.33/0.1666	86.13/0.7226	88.70/0.7740	92.99/0.8597	93.25/0.8650
T_{S_6}	72.50/0.4500	74.90/0.4980	90.30/0.8060	88.24/0.7649	93.06/0.8612
T_{S_7}	63.33/0.2666	78.00/0.5600	78.40/0.5680	94.41/0.8882	95.94/0.9188
T_{S_8}	55.83/0.1166	79.85/0.5970	91.30/0.8260	93.00/0.8600	94.82/0.8964
T_{S_9}	70.00/0.4000	79.90/0.5980	88.40/0.7680	86.50/0.7300	91.75/0.8350
Avg	66.85/0.3437	75.80/0.5159	84.70/0.6936	86.07/0.7213	92.56/0.8513
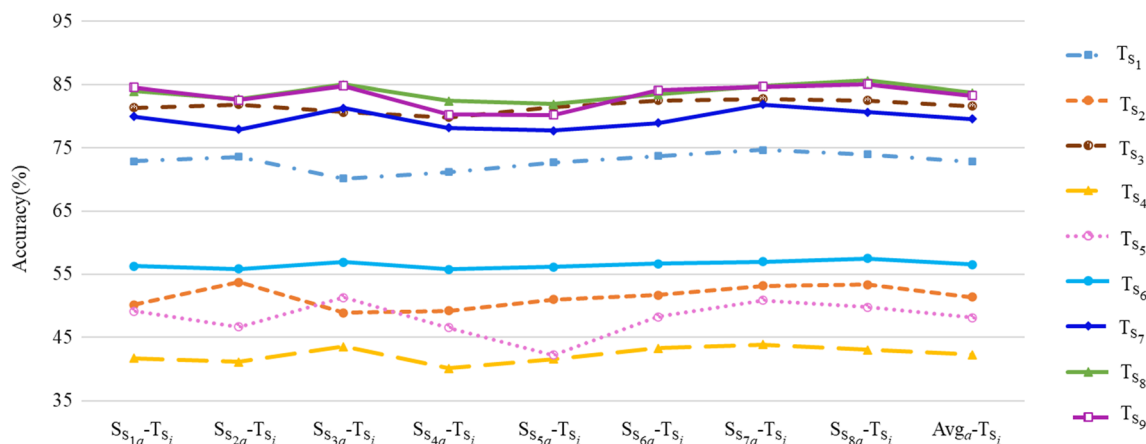
Fig. 14 Classification accuracies of different single-source to target transfers on dataset 2

the decision boundary of the target domain, and multi-source to target transfers greatly improve the performance of the model.

5.1.2 Dataset 2: BCI IV 2a

To demonstrate the effectiveness of DAMSDAF in multiple class, the classification accuracies of single-source to target transfers and the average accuracy of each target subject on dataset 2 are shown in Fig. 14. The meaning of each mathematical symbol in this subsection is consistent with those in dataset 1. We can see that the classification results of single-source to target transfers are different. In addition, the highest accuracy of single-source to target transfers is higher than the average accuracy; therefore, it is also necessary to optimize the source domain for multiple tasks. Figure 15 shows the classification accuracies of each target subject through DAMSDAF on dataset 2. The tendency in Fig. 15 is relatively

consistent with that in Fig. 12, and the comparison results of Figs. 14 and 15 are consistent with those of Figs. 11 and 12, demonstrating that the proposed method is effective in multi-class classification. Figure 16 shows the confusion matrix of the DAMSDAF method on dataset 2. FT represents the imaginary foot task, and T represents the imaginary tongue task. The classification accuracies of LH, RH, FT and T are 71.11%, 70%, 67.78% and 68.89%, respectively, indicating that the accuracy of task T is slightly worse than that of LH and RH tasks, while the FT task is relatively indistinguishable.

Table 5 shows the classification results of various domain adaptation methods on dataset 2. $DAMSDAF_{5S}$ represents the results of 5 target subjects. It can be seen in Table 5 that DAMSDAF achieved the highest results, and the gaps in DAMSDAF and DASSDAF, W/C, ASFA, MFTL-TSK and SCSN-MMD are 2.85%, 15.67%, 12.09%, 0.95% and 1.01%, respectively. W/C selected one of the most relevant source subjects from multiple source domains, and realized

Fig. 15 Variation in classification accuracy with an increasing number of source domains on dataset 2
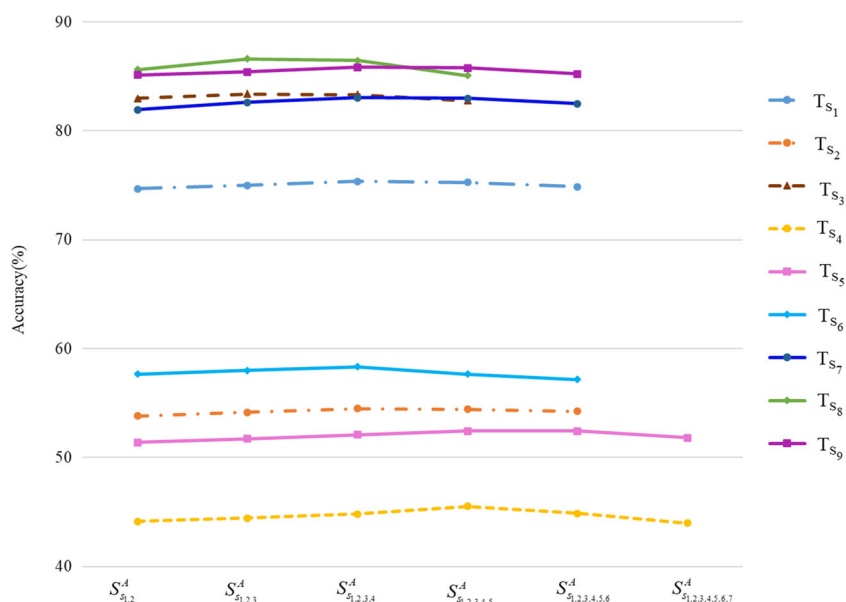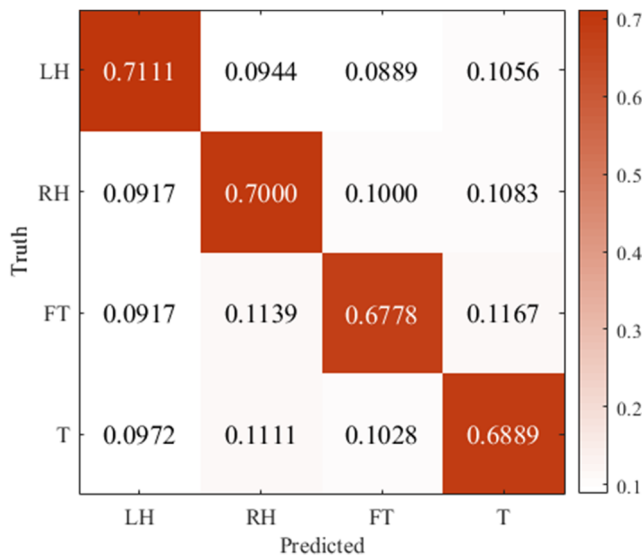
Fig. 16 Confusion matrix obtained with the average of the optimal multi-source to target transfers on dataset 2

domain adaptation for one pair of source and target domains by a gradient reversal layer, while DAMSDAF performed the multi-source to target transfers by using MMD to align each pair of ASD and target domain simultaneously, and the results obtained by DAMSDAF were higher than W/C. ASFA used sample reweighting to align the classifier output between auxiliary and target classifiers, and the classification accuracy was higher than W/C by 3.58%. MFTL-TSK used information from other source subjects and a small amount of MI-EEG data from target subject to conduct the distribution alignment and achieved a great improvement in classification accuracy. SCSN-MMD successively selected 5 subjects with the highest data quality (S_1, S_3, S_7, S_8 and S_9) from dataset 2 as the target subject for evaluation, and the data distribution among subjects might not be the most similar. DAMSDAF enhanced the similarity between the

source and target domains by pre-aligning the source domain, and the average result of DAMSDAF outperformed SCSN-MMD on these 5 subjects. Furthermore, Table 5 also shows kappa values on dataset 2. There is a large difference among subjects. S8 obtained the highest value of 0.8213, while S4 obtained the lowest value of 0.2733, but DAMSDAF showed a much-improved kappa value for all subjects compared with that of the related methods. Overall, the results obtained by DAMSDAF were higher than those obtained by relevant transfer learning methods, indicating the importance of multi-source domains transfer learning by combining source domain alignment and multi-source domain adaptation.

5.2 Cross-dataset transfer learning

To evaluate the performance of DAMSDAF on cross-dataset transfer learning, Fig. 17 shows the classification results of single-source to target transfers and multi-source to target transfers on dataset 3. T_{S_1} represents the one target subject in dataset 3, and $S_{S_h}^A$-T_{S_1} represents the knowledge transfer from the hth ASD of dataset 1 to the target domain. The blue bar graph represents the results and average result of single-source to target transfers, and the orange bar graph represents the results of multi-source to target transfers. DAMSDAF obtained the highest classification accuracy with three source domains, demonstrating that DAMSDAF can be used for knowledge transfer from multiple source domains to target domain with greater data distribution differences. Figure 18 shows the confusion matrix of the DAMSDAF method on dataset 3. The classification accuracies of LH and RH are 89.14% and 90.00%, respectively, which achieved near-unanimous results, indicating that DAMSDAF also contributes greatly to cross-dataset transfer learning. However, comparing Figs. 12, 15 and 17, it can be obtained that the

Table 5 Comparison of the accuracies (%) and kappa values (accuracy/kappa value) of DAMSDAF with different transfer learning methods on dataset 2

| Transfer learning methods							
Target subject	W/C [32]	ASFA [14]	MFTL-TSK [17]	DASSDAF	DAMSDAF	SCSN-MMD [37]	DAMSDAF$_{5S}$
TS_1	70.00/0.6000	72.99/0.6399	79.00/0.7200	72.84/0.6379	75.35/0.6713	76.80/0.6907	75.35/0.6713
TS_2	38.00/0.1700	28.19/0.0425	52.00/0.3600	51.40/0.3520	54.51/0.3935	–	–
TS_3	76.00/0.6900	86.25/0.8167	85.75/0.8100	81.63/0.7551	83.40/0.7787	80.50/0.7400	83.40/0.7787
TS_4	39.00/0.1900	42.95/0.2393	58.75/0.4500	42.26/0.2302	45.50/0.2733	–	–
TS_5	31.00/0.0700	36.01/0.1468	45.25/0.2700	48.10/0.3080	52.43/0.3657	–	–
TS_6	38.00/0.1800	41.60/0.2213	49.00/0.3200	56.53/0.4204	58.33/0.4444	–	–
TS_7	70.00/0.5900	69.51/0.5935	82.00/0.7600	79.57/0.7277	83.06/0.7741	80.60/0.7413	83.06/0.7741
TS_8	63.00/0.5000	78.78/0.7171	82.00/0.7600	83.74/0.7832	86.60/0.8213	82.40/0.7653	86.60/0.8213
TS_9	59.00/0.4500	59.90/0.4653	80.50/0.7400	83.28/0.7771	85.83/0.8111	88.90/0.8520	85.83/0.8111
Avg	53.78/0.3822	57.35/0.4314	68.50/0.5767	66.60/0.5546	69.45/0.5926	81.84/0.7579	82.85/0.7713
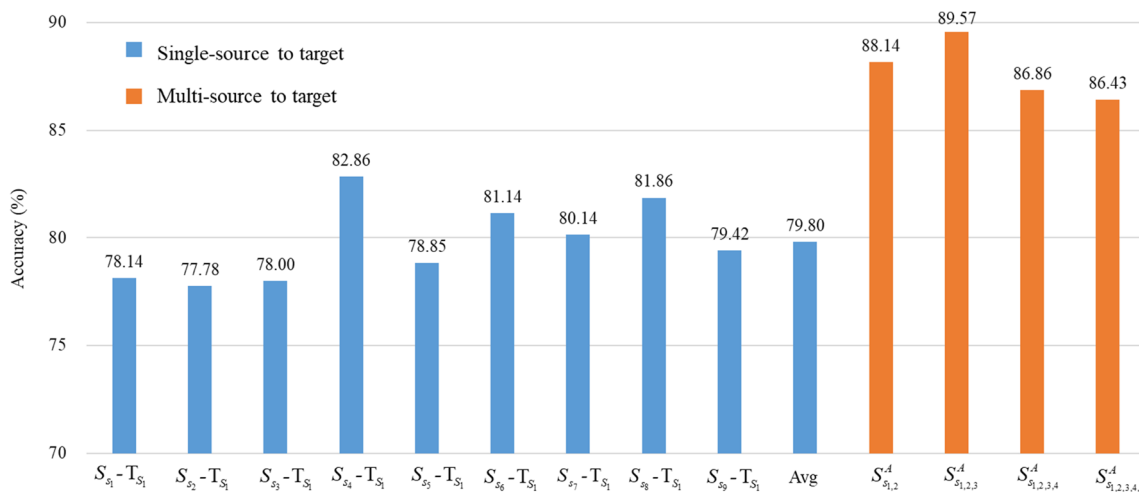
Fig. 17 Classification accuracies of single-source and multi-source to target transfers on dataset 3

classification accuracy of the cross-dataset decreases faster than that of inter-subject, which might be because adding more weakly related source domains, reveals that the correlation among domains is still the basis of transfer learning.

5.3 Statistical analysis

To further analyze the performance improvements of DAMSDAF over each compared transfer learning method on classification results, paired t-test was performed to detect whether there is a significant difference when they are applied to recognize the MI-EEG in this paper.

First, the Lilliefors test was used to verify whether the classification results produced by the proposed methods and other compared methods come from a normal distribution. In this experiment, the *lillietest* function of MATLAB was used to test the normal distribution. The test results of datasets 1

Fig. 18 Confusion matrix obtained with the optimal multi-source to target transfers on dataset 3

and 2 are displayed in Table 6, respectively. The output results include the hypothesis test result h and the p value, where h is returned as a logical value, and p is returned as a scalar value in the range (0.0, 1.0). The output results of all approaches are $h = 0$ and $p > 0.05$, which means that the null hypothesis that the data from a normal distribution cannot be rejected. Hence, the classification results of DAMSDAF and the state-of-the-art methods all fit the normal distribution except for MFTL-TSK.

After the normal distribution of all methods was examined, we performed the paired t-test by using the MATLAB *ttest* function. Assume that two samples were chosen from the abovementioned methods, and they had the same sample size θ, $x_{1i} \in$ sample1, $x_{2i} \in$ sample2, $i = 1, 2, \cdots, \theta$, and then the test statistic t was calculated by

$$t = \frac{\overline{x} - \mu}{\sqrt{\frac{\sum_{i=1}^{\theta}\left(x_i - \overline{x}\right)}{\theta - 1}} / \sqrt{\theta}} \tag{19}$$

$$\overline{x} = \frac{\sum_{i=1}^{\theta} x_i}{\theta} \tag{20}$$

where $x_i = x_{1i} - x_{2i}$ is the difference between all pairs and $\overline{x}$ is the average value. In this section, we want to test whether the average of the difference is significantly different, and the constant $\mu = 0$.

Defining the null hypothesis is H_0: the difference between the paired samples has a mean of zero; the alternative hypothesis is H_1: the classification results of each pairwise comparison have unequal means. The significance level can be set as $\alpha = 0.05$. The decision rule is to reject H_0 if

$$p = P\{t > t_\alpha(n-1)\} \leq 0.05 \tag{21}$$

The paired t-test results on the classification accuracies of datasets 1 and 2 are shown in Table 7, respectively, where the
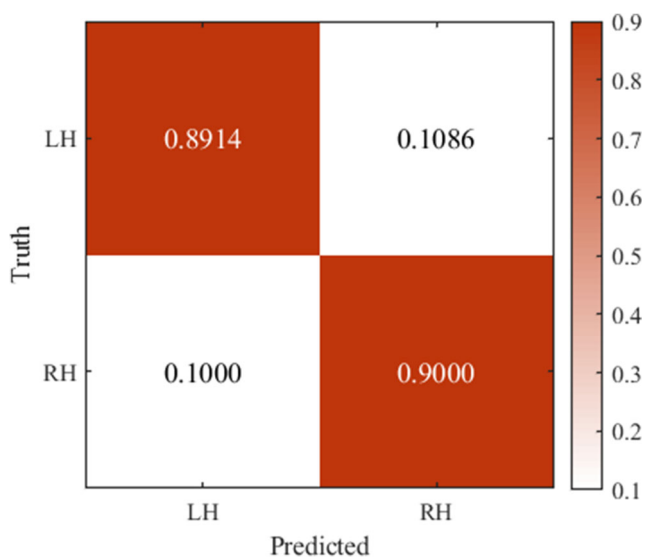
Table 6 Results of the normal distribution test on datasets 1 and 2

Type	Dataset	Group	Mean	p value	h
Normal distribution test	Dataset 1	SIITAL$_{fbcsp}$ [11]	75.80	0.5	0
		SBTACSP [18]	66.85	0.5	0
		MF-BLDA [15]	84.70	0.1587	0
		DASSDAF	86.07	0.3941	0
		DAMSDAF	92.56	0.5	0
	Dataset 1	W/C [32]	53.78	0.1082	0
		ASFA [14]	57.35	0.3629	0
		MFTL-TSK [17]	68.50	0.0224	1
		SCSN-MMD [37]	81.84	0.3893	0
		DASSDAF	66.60	0.3522	0
		DAMSDAF	69.45	0.1443	0

6 p values are smaller than 0.05 in datasets 1, except for the p value between MF-BLDA and DASSDAF, and 5 of the 6 p values are smaller than 0.05 in datasets 2, except for the p value between SCSN-MMD and DASMDAF. Hence, the null hypothesis H_0 is rejected at the 0.05 significance level, indicating that the differences between DAMSDAF and each other method are statistically significant and showing the advantage of the proposed method in the recognition of MI-EEG. In addition, DAMSDAF significantly outperformed DASSDAF on datasets 1 and 2 by observing the last column in Table 7, suggesting that the transfer effectiveness of multi-source to target transfers is superior to that of single-source to target transfers.

6 Discussion

Considering the small data size and individual variabilities of MI-EEG, a multi-source domain adaptation framework based on dual alignment was proposed to explore multiple source domain transfer learning in this paper. The performance of DAMSDAF was evaluated by 3 MI datasets and compared with the relevant research results, as shown in Tables 3, 4 and 5. DAMSDAF obtained average recognition rates of 92.56% and 69.45% and average kappa values of 0.8513 and 0.5926 on datasets 1 and 2, respectively, which were higher than those of the compared methods. In addition, DAMSDAF achieved an 89.57% recognition rate on dataset 3, which is higher than DASSDAF. It is

confirmed that DAMSDAF is effective and universal for inter-subject and cross-dataset transfer learning.

Figures 11 and 14 show the classification results of DASSDAF on datasets 1 and 2, respectively. The average accuracy of DASSDAF was not the optimal method for multi-source to target transfers. DAMSDAF can select multiple source domains that provide auxiliary information to the target domain and recognize the MI-EEG of the target domain through MBDN and weighted prediction to obtain the best transfer effect. Figures 12 and 15 show the classification results of DAMSDAF with multi-source domains. Interestingly, comparing Figs. 11 and 12 and Figs. 14 and 15, with the increase in the number of source domains, the accuracy of multi-source to target transfers decreases, and the results are lower than that those of the best single-source to target transfers on the same target subject, demonstrating that more source domains are used for multi-source domain adaptation, which is not always better. However, the sequential selection algorithm can solve this phenomenon well by selecting the most appropriate number of source domains for transfer learning. Enchantingly, the multiple class MI tasks need more knowledge transfer of the source domain to achieve the best classification accuracy by comparing Figs. 12 and 15, which might be because the domain adaptation of multiple class MI tasks was more complicated than two class MI tasks. The produced $\mathcal{L}_{MMD}$ had a large impact on the MBDN in the late training process, which might cause the whole network training to be unstable; therefore, the multiple class MI tasks need more source domains to train the MBDN. In addition, by

Table 7 p-Values of paired t-test results on the classification accuracies of datasets 1 and 2

Dataset 1	Methods	SIITAL$_{fbcsp}$ [11]	SBTACSP [18]	MF-BLDA [15]	DASSDAF
	DASSDAF	7.1240e-04	0.0023	0.6215	–
	DAMSDAF	4.8017e-04	2.8008e-05	0.0059	0.0259
Dataset 2	Methods	W/C [32]	ASFA [14]	SCSN-MMD [37]	DASSDAF
	DASSDAF	0.0013	0.0264	–	–
	DAMSDAF$_{/5S}$	3.8777e-04	0.0082	0.5080	5.8305e-06

observing the number of source domains used on target domains in Figs. 12 and 15, it is difficult to distinguish MI tasks that might need more source domains, while the space for improvement is considerable, such as T_{S_2} and T_{S_3} in dataset 1 and T_{S_4} and T_{S_5} in dataset 2, indicating that DAMSDAF is a universal framework that can maximize the auxiliary of the target domain.

To determine the influence of ASD, MMD and weighted average (WA) on the proposed method, we conducted experiments on DAMSDAF without one of these three strategies, and the data division and accuracy acquisition methods are consistent with DAMSDAF. The average classification results of the target domains on three datasets with different strategies are shown in Fig. 19. DAMSDAF$_{ASD}$ represents the obtained results without considering the ASD, i.e., the samples in the source subject are directly transferred to D_{T_T} as the source domain without considering weight assignment. **DAMSDAF$_{\mathcal{L}_{MMD}}$** represents the obtained results without considering the MMD, i.e., each pair of ASD and target domain is not aligned after extracting the domain-specific features. DAMSDAF$_{WA}$ represents the obtained results without considering the WA, i.e., the output of the sub-classifiers is not weighted while directly averaging the output of M classifiers as the final result. The blue, green and orange lines represent the results of four different DAMSDAFs on datasets 1, 2 and 3, respectively. It can be seen that the results of DAMSDAF$_{\mathcal{L}_{MMD}}$ are the lowest among the four methods on 3 datasets, indicating that the feature-level domain adaptation has the greatest influence on the transfer learning of multi-source domains. The gaps in DAMSDAF$_{WA}$ and DAMSDAF are 1.31% and 0.38% on datasets 1 and 2, respectively, demonstrating WA has a slightly greater impact on dataset 1 than dataset 2. The cause of this phenomenon might be that each single-source domain in dataset 1 has a great difference from the auxiliary for the target domain, and the most relevant source domains can perform better with WA. Moreover, the gaps of DAMSDAF$_{ASD}$ and DAMSDAF are 1.93% and 2.09% on datasets 1 and 2, respectively,
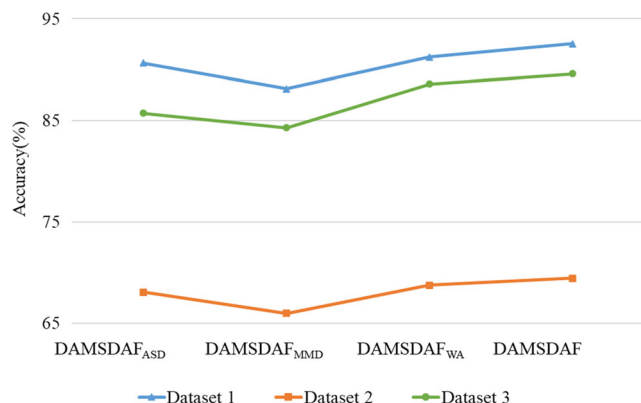
illustrating that ASD has a greater influence on transfer learning of inter-subject with multiple MI tasks, which might be because the relatively large $\mathcal{L}_{MMD}$, which is generated by the unaligned source domain and target in the specific feature space, affects the performance of MBDN. In addition, the results of DAMSDAF$_{WA}$ and DAMSDAF$_{ASD}$ on dataset 3 are different from the results of DAMSDAF by 2.43% and 3.86%, respectively, which are larger than the difference on datasets 1 and 2, confirming WA and ASD have less influence on inter-subject, but more influence on cross-dataset transfer learning. The main reason is that there are large data distribution discrepancies between datasets 1 and 3. Therefore, reducing the differences in inter-subject transfer learning, especially in cross-dataset transfer learning is still a great challenge for future consideration. Overall, ASD, MMD and WA all have a certain influence on the proposed method. DAMSDAF achieved preferable results on 3 datasets by combining three strategies, which suggests that DAMSDAF is feasible and effective for multi-source to target transfers in MI-EEG classification.

Due to the specificity of EEG, the information obtained by different channels with great discrepancy, and the influence of different channels on the classification accuracy are explored in this work. The channel-wise results obtained for the proposed method on three datasets are shown in Fig. 20, where C3-Cz denotes combining the information of the C3 and Cz channels, C4-Cz denotes combining the information of the C4 and Cz channels, C3-C4 denotes combining the information of the C3 and C4 channels, C3-Cz-C4 denotes combining the information of the C3, Cz and C4 channels. The results of DAMSDAF all achieved the best performance by combining the C3, Cz and C4 channels on three datasets, while that of DAMSDAF achieved the lowest performance by using the Cz channel on datasets 1 and 3. This is because the Cz channel has very little information related to the MI of the left or right hands, while the positions of C3 and C4 channels are the main MI brain areas for left and right hands, which provide EEG signals that can reflect brain activity. However, it can be seen

Fig. 19 Influence on the performance of DAMSDAF without considering ASD, MMD or WA
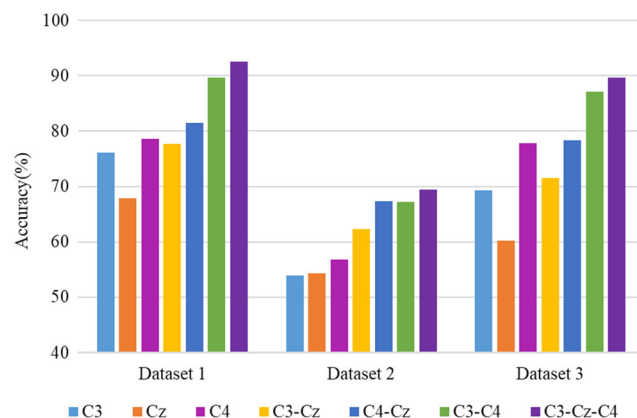
Fig. 20 Channel-wise accuracies obtained for DAMSDAF on three datasets

in dataset 2 that the classification accuracies of Cz and C4-Cz are slightly higher than those of C3 and C4-C3, and the gaps are 0.30% and 0.12%, respectively, indicating that Cz is beneficial for recognizing the MI of the foot and tongue. In addition, as seen in dataset 1, the classification accuracy of C4 is higher than that of C3 by 2.50%, and the classification accuracy of C3-C4 is higher than that of C3 and C4 by 13.44% and 10.94%, respectively, indicating that C4 acquires the most relevant information during MI, and combining C3 and C4 can greatly improve the results. Furthermore, the classification accuracies of C4-Cz-C3, C4-Cz and C3-Cz are higher than those of C4-C3, C4 and C3 by 2.86%, 2.81% and 1.56%, respectively, indicating that combining Cz can improve the classification result. The findings from dataset 1 can also be obtained on the other two datasets. The channel-wise classification results suggest that using more channels will achieve better MI-EEG classification results.

Based on the above discussion and results analysis, the main advantages of this paper are summarized into three points. First, the data distribution between different source and target domains is aligned based on NMI, which provides sufficient and reliable training data for the classifier. Second, a multi-branch deep network is designed based on MMD, which aligns the feature distributions of each pair of source and target domains by learning the outputs of several sub-neural networks to strengthen the model training. Finally, a sequential selection algorithm is applied to select optimal multiple source domains, which proves that multi-source transfer learning is superior to single-source transfer learning on three datasets. As seen in Tables 4 and 5 the classification accuracies of DAMSDAF outperform the related works. In addition, the weighted alignment and domain adaptation can easily be embedded into the real-time BCI as predestinate strategies, which has important significance in a clinical scenario. This work lays a solid foundation for the application of domain adaptation in MI-EEG classification, and will promote the more extensive integration of MI-EEG recognition with domain adaptation technology and deep learning.

However, there are several limitations in this work, which will be addressed in our future research:

1. DAMSDAF copes well with the MI-EEG classification on 3 EEG channels (C3, Cz and C4), ignoring the information carried by other channels. For complex MI tasks, the activated brain regions may overlap, and fewer channels carry less comprehensive information, leading to suboptimal results. More experiments will be carried out in future work to further verify the universality of DAMSDAF.
2. The current DAMSDAF can be used in multi-source domains transfer learning; however, it is time-consuming in

this work. This is due to the server configuration and the limited operating speed. We will equip the laboratory with higher-performance experimental equipment in the future.
3. The datasets used in this paper were derived from public databases, and the samples were limited to normal subjects. This is due to the influences of COVID-19, affecting the data collection of stroke patients, which was carried out as early as possible and applied to demonstrate the effectiveness of the proposed method.

7 Conclusions

Aiming at the individual difference characteristics of complex MI-EEG signals in multi-source domains transfer learning, the current research focuses on aligning the data distribution of the source and target domains. This paper proposes a novel method called DAMSDAF, which consists of three steps. First, calculate the NMI between the time-frequency spectrum images of the source domain and informative samples to align the data distribution of the source domain, making the decision boundary of ASD closer to that of the target domain, which is beneficial to model training. Second, design an MBDN with MMD, which aligns the domain-specific feature distribution of each pair of ASD and target domain simultaneously by learning multiple domain invariant representations, and the output of multiple classifiers is a weighted average that can reasonably utilize the knowledge transfer from multiple source domains to the target domain with different degrees of correlation. Third, select the optimal multi-source domains for transfer learning via the sequential selection algorithm to ensure the best transfer effect, enabling transfer learning of multi-source domains in BCI. Experiments on three MI-EEG datasets of inter-subject and cross-dataset transfer learning demonstrated that DAMSDAF outperformed the state-of-the-art transfer learning methods, indicating that the dual alignment can make the data distribution of source domains more similar to the target domain and verifying the validity of DAMSDAF. It is expected to promote the application of BCI technology and transfer learning in rehabilitation engineering. The main drawback of this work is that we did not fully consider the applicability of the shared network, and training a domain common network to improve the proposed framework is a potential research direction.

This paper briefly presents the application of domain adaptation in MI-EEG recognition, but other new deep learning methods, especially the attention mechanism and transformer, are not considered. The attention mechanism draws on the method of human attention thinking, considers different weights for each input element and focuses more on the parts that are similar to the point of interest while suppressing useless information. DAMSDAF combined with the attention

mechanism may enhance the information of samples with high NMI values by weighting the features similar to the target domain. The transformer achieves efficient parallelization and another very important innovation is the use of positional encoding, which encodes different positions to a certain extent. DAMSDAF successfully realizes multi-source domain transfer learning but is time-consuming, while transformers may be used as common feature extractor to extract the features of the multiple source domains and target domain and improve model efficiency through parallel computing. In addition, the proposed model has high application value and can be used to develop intelligent wearable devices to help stroke patients recover motor control. In the future, we plan to develop a practical rehabilitation training system that can use a wearable cap to acquire the MI-EEG signals of patients, input them into our proposed model to recognize MI-EEG, and transmit the results to the wearable robotic arm to help patients undergo rehabilitation training. Furthermore, the proposed model can also be used in the development of virtual reality games and smart-home devices, thereby expanding the application prospects of BCI.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Loh HW, Ooi CP, Aydemir E et al (2021) Decision support system for major depression detection using spectrogram and convolution neural network with EEG signals. Expert Syst 39(3):e12443

2. Tuncer T, Dogan S, Maygin M et al (2022) Tetromino pattern based accurate EEG emotion classification model. Artif Intell Med 123:102210

3. Dogan A, Akay M, Barua PD, Baygin M, Dogan S, Tuncer T, Dogru AH, Acharya UR (2021) PrimePatNet87: prime pattern and tunable q-factor wavelet transform techniques for automated accurate EEG emotion recognition. Comput Biol Med 138:104867

4. Barua PD, Dogan S, Tuncer T, Baygin M, Acharya UR (2021) Novel automated PD detection system using aspirin pattern with EEG signals[J]. Comput Biol Med 137:104841

5. Pappalettera C, Miraglia F, Cotelli M, Rossini PM, Vecchio F (2022) Analysis of complexity in the EEG activity of Parkinson's disease patients by means of approximate entropy. Geroscience 44(3):1599–1607

6. Zabcikova M, Koudelkova Z, Jasek R, Lorenzo Navarro JJ (2022) Recent advances and current trends in brain-computer interface research and their applications. Int J Dev Neurosci 82(2):107–123

7. Simanto S, Mathias B (2020) Intra- and inter-subject variability in EEG-based sensorimotor brain computer Interface: a review. Front Comput Neurosci 13. https://doi.org/10.3389/fncom.2019.00087

8. Tiwari S, Goel S, Bhardwaj A (2022) MIDNN- a classification approach for the EEG based motor imagery tasks using deep neural network. Appl Intell 52(2):4824–4843

9. Wu DR, Xu YF, Lu BL (2022) Transfer learning for EEG-based brain-computer interfaces: a review of Progress made since 2016. IEEE Transact Cogn Dev Syst 14(1):4–19

10. Xu L, Xu M, Ke Y, An X, Liu S, Ming D (2020) Cross-dataset variability problem in EEG decoding with deep learning. Front Hum Neurosci 14(103)

11. Ibrahim H, Abbas K, Imali T H, et al (2017) Informative instance transfer learning with subject specific frequency responses for motor imagery brain computer interface. 2017 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 252-257

12. Ibrahim H, Abbas K, Imali H et al (2018) Multiclass informative instance transfer learning framework for motor imagery-based brain-computer Interface. Comput Intel Neurosci 2018. https://doi.org/10.1155/2018/6323414

13. Wu DR (2017) Online and offline domain adaptation for reducing BCI calibration effort. IEEE Transact Human Machine Syst 47(4):550–563

14. Kun X, Deng L, Wlodzislaw D et al (2022) Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces. IEEE Trans Biomed Eng PP. https://doi.org/10.1109/TBME.2022.3168570

15. Zhang Y, Chen W, Lin CL, Pei Z, Chen J, Chen Z (2021) Boosting-LDA algriothm with multi-domain feature fusion for motor imagery EEG decoding. Biomed Signal Proc Contr 70:102983

16. Zhang Y, Li H, Dong H, Dai Z, Chen X, Li Z (2022) Transfer learning algorithm Design for Feature Transfer Problem in motor imagery brain-computer Interface[J]. China Commun 19(2):39–46

17. Liang Y, Ma Y (2020) Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning. Biomed Signal Proc Contr 62:102101

18. Zhang XX, She QS, Chen Y, Kong W, Mei C (2021) Sub-band target alignment common spatial pattern in brain-computer Interface. Comput Methods Prog Biomed 207:106150

19. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359

20. Day O, Khoshgoftaar TM (2017) A survey on heterogeneous transfer learning. J Big Data 4:29

21. Lan Z, Olga S, Wang L et al (2019) Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public data sets. IEEE Transact Cogn Dev Syst 11(1):85–94

22. Wang J, Lan C, Liu C, et al (2021) Generalizing to unseen domains: a survey on domain generalization. arXiv preprint arXiv:2103.03097

23. Chai X, Wang Q, Zhao Y, Liu X, Bai O, Li Y (2016) Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. Comput Biol Med 79:205–214

24. Cai Y, She Q, Ji J, Ma Y, Zhang J, Zhang Y (2022) Motor imagery EEG decoding using manifold embedded transfer learning. J Neurosci Met 370:109489

25. Zhu L, Yang JT, Ding WP, Zhu J, Xu P, Ying N, Zhang J (2021) Multi-source fusion domain adaptation using resting-state knowledge for motor imagery classification tasks. IEEE Sensors J 21(19):21772–21781

26. Jiang Q, Zhang Y, Zheng K (2022) Motor imagery classification via kernel-based domain adaptation on an SPD manifold. Brain Sci 12(5):656

27. Liu BC, Chen X, Li X, Wang Y, Gao X, Gao S (2021) Align and pool for EEG headset domain adaptation (ALPHA) to facilitate dry electrode based SSVEP-BCI. IEEE Trans Biomed Eng 69:795–806. https://doi.org/10.1109/TBME.2021.3105331

28. Peterson V, Nieto N, Wyser D, Lambercy O, Gassert R, Milone DH, Spies RD (2022) Transfer learning based on optimal transport for motor imagery brain-computer interfaces. IEEE Trans Biomed Eng 69(2):807–817

29. Shen M, Zou B, Li XH, Zheng Y, Li L, Zhang L (2021) Multi-source signal alignment and efficient multi-dimensional feature classification in the application of EEG-based subject-independent drowsiness detection. Biomed Sign Proc Cont 70:103023. https://doi.org/10.1016/j.bspc.2021.103023

30. Ko W, Jeon E, Jeong S, Phyo J, Suk HI (2021) A survey on deep learning-based short/zero-calibration approaches for EEG-based brain-computer interfaces. Front Hum Neurosci 15. https://doi.org/10.3389/fnhum.2021.643386

31. Long SF, Wang SS, Zhao X et al (2022) Cross-domain feature enhancement for unsupervised domain adaptation. Appl Intell. https://doi.org/10.1007/s10489-022-03306-9

32. Jeon E, Ko W, Suk H (2019) Domain adaptation with source selection for motor-imagery based BCI. In: 2019 7Th international winter conference on brain-computer Interface (BCI). IEEE, pp 134-137

33. Jeon E, Ko W, Yoon JS et al (2019) Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI. IEEE Transact Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3100583

34. Chen PY, Gao ZK, Yin MM (2021) Multiattention adaptation network for motor imagery recognition. IEEE Transact Syst, Man, Cybern: Syst. https://doi.org/10.1109/TSMC.2021.3114145

35. Hong X, Zheng Q, Liu L, Chen P, Ma K, Gao Z, Zheng Y (2021) Dynamic joint domain adaptation network for motor imagery classification. IEEE Transact Neural Syst Rehabili Engin 29:556–565

36. Zhao H, Zheng Q, Ma K, Li H, Zheng Y (2021) Deep representation-based domain adaptation for nonstationary EEG classification. IEEE Transact Neural Netw Learn Syst 32(2):535–545

37. Wei X, Ortega P, Faisal A A (2021) Inter-subject deep transfer learning for motor imagery EEG decoding. International IEEE EMBS conference on neural engineering. IEEE, pp 21-24

38. Zheng MM (2021) Yang BH (2021) a deep neural network with subdomain adaptation for motor imagery brain-computer interface. Med Eng Phys 96:29–40

39. Ben D S, Blitzer J, Crammer K, Kulesza, et al (2010) A theory of learning from different domains. Mach Learn 79(1–2): 151–175

40. Sun SL, Shi HL, Wu YB (2015) A survey of multi-source domain adaptation. Inform Fusion 24:84–92

41. Leeb R, Brunner C, Miiller-Putz G [online]. BCI competition IV, 2008. Available: http://www.bbci.de/competition/iv/, Accessed on 27 April 2021

42. Brunner C, Leeb R, Miiller-Putz G [online]. BCI competition IV, 2008. Available: http://www.bbci.de/competition/iv/, Accessed on 27 April 2021

43. Gert Pfurtscheller, Alois Schlögl [Online]. BCI Competition II, 2002. Available online: https://www.bbci.de/competition/ii/#Data sets, Accessed on 27 April 2021

44. Witten IH, Frank E (2011) Data mining: practical machine learning tools and techniques. ACM SIGMOD Rec 31(1):76–77

45. Ganin Y, Lempitsky V (2015) Unsupervised domain adaptation by backpropagation. Proc Machine Learn Res 37:1180–1189