



Multi-Source and Multi-Representation Adaptation for Cross-Domain Electroencephalography Emotion Recognition

Jiangsheng Cao¹, Xueqin He¹, Chenhui Yang^{1*}, Sifang Chen^{2*}, Zhangyu Li² and Zhanxiang Wang^{3,4}

¹ School of Informatics, Xiamen University, Xiamen, China, ² Department of Neurosurgery, The First Affiliated Hospital of Xiamen University, Xiamen, China, ³ Department of Neurosurgery, Xiamen Key Laboratory of Brain Center, The First Affiliated Hospital of Xiamen University, Xiamen, China, ⁴ Department of Neuroscience, Institute of Neurosurgery, School of Medicine, Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:

Yuanpeng Zhang,
Nantong University, China

Reviewed by:

Xin Huang,
Jiangxi Normal University, China
Guokai Zhang,
Tongji University, China

*Correspondence:

Chenhui Yang
chyang@xmu.edu.cn
Sifang Chen
cfsong143@aliyun.com

Specialty section:

This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: 05 November 2021

Accepted: 21 December 2021

Published: 13 January 2022

Citation:

Cao J, He X, Yang C, Chen S, Li Z
and Wang Z (2022) Multi-Source
and Multi-Representation Adaptation
for Cross-Domain
Electroencephalography Emotion
Recognition.
Front. Psychol. 12:809459.
doi: 10.3389/fpsyg.2021.809459

Due to the non-invasiveness and high precision of electroencephalography (EEG), the combination of EEG and artificial intelligence (AI) is often used for emotion recognition. However, the internal differences in EEG data have become an obstacle to classification accuracy. To solve this problem, considering labeled data from similar nature but different domains, domain adaptation usually provides an attractive option. Most of the existing researches aggregate the EEG data from different subjects and sessions as a source domain, which ignores the assumption that the source has a certain marginal distribution. Moreover, existing methods often only align the representation distributions extracted from a single structure, and may only contain partial information. Therefore, we propose the multi-source and multi-representation adaptation (MSMRA) for cross-domain EEG emotion recognition, which divides the EEG data from different subjects and sessions into multiple domains and aligns the distribution of multiple representations extracted from a hybrid structure. Two datasets, i.e., SEED and SEED IV, are used to validate the proposed method in cross-session and cross-subject transfer scenarios, experimental results demonstrate the superior performance of our model to state-of-the-art models in most settings.

Keywords: EEG, emotion recognition, domain adaption, deep learning, affective computing, SEED

INTRODUCTION

Emotion is a physiological state of humans, which appears when people are stimulated by external or their own factors. Emotion is the basis of human daily life and work and plays an important role in human psychological development, interpersonal communication, rational decision-making, and cognition. Correct recognition of emotions is of great significance in the fields of education, medical treatment, psychology, cognitive science, and artificial intelligence. For example, in the medical field, emotion recognition can help doctors diagnose and treat patients with expression

disorders; in the education field, through emotion recognition, teachers can quickly find and encourage students with poor emotions to create a more positive learning environment.

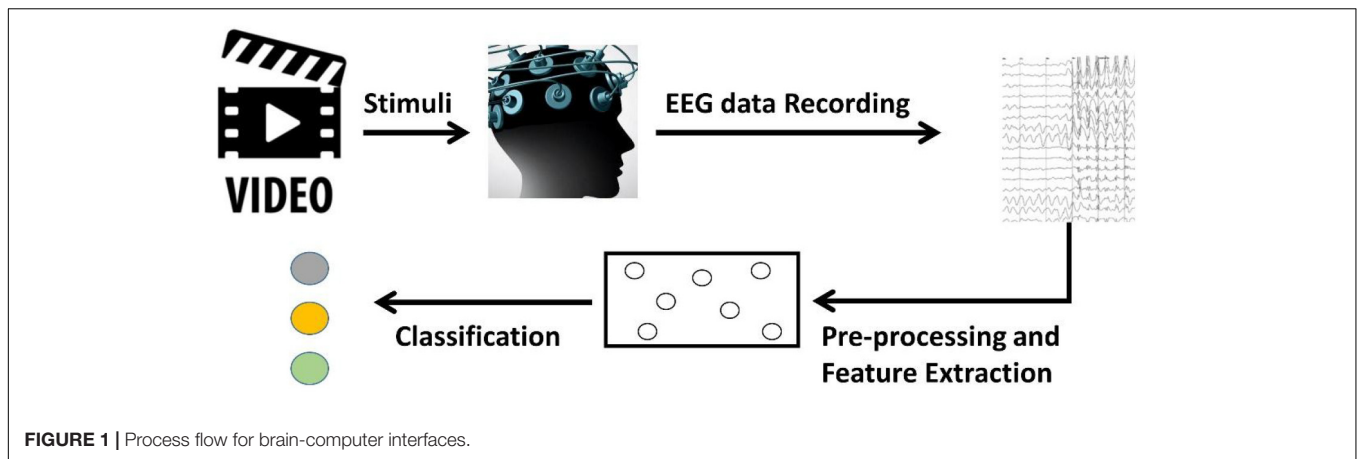
Data sources used for emotion recognition can be roughly divided into two categories: non-physiological signals and physiological signals. Many scholars have carried out researches on emotion recognition using non-physiological signals such as gestures, facial expressions, eye expressions, and voice. Said and Barr (2021) proposed a face-sensitive convolutional neural network (FS-CNN) to detect human emotion. Trigeorgis et al. (2016) utilized a convolutional recurrent model based on the raw signal to achieve end-to-end spontaneous emotion prediction. In addition, some scholars (Camurri et al., 2003; Durupinar et al., 2016; Senecal et al., 2016) try to identify emotions from dance movements. These methods are limited to specific dance moves and lack practical significance. Due to the easy camouflage of non-physiological signals, those non-physiological signal-based emotion recognition methods are unstable, and the recognition effect is easily affected by subjective factors. Compared with non-physiological signals, physiological signals [such as blood pressure (BVP), electroencephalogram (EEG), electrooculogram (EOG), electrocardiography (ECG), electromyogram (EMG), etc.] are spontaneously generated by the human body and can truly reflect the emotional state of humans, which has high reliability. Therefore, scholars have shifted the focus of research to emotion recognition using physiological signals. Among physiological signals, EEG, which has the characteristics of non-subjectivity, real-time difference, and rich information, has been widely used in the field of emotion recognition. Studies have shown that EEG played an important role in the research of human emotion and that regional brain activity was closely related to some emotional states (Niemic, 2004).

With the continuous development of artificial intelligence technology, artificial intelligence has also shown a broader prospect in the field of emotion recognition based on EEG. Employing emerging intelligent technologies like machine learning and computer vision to realize EEG-based quantitative analysis and recognition of emotions has become a research hotspot. Existing EEG-based emotion recognition methods can be roughly divided into two categories: traditional machine learning-based methods and deep learning-based methods. Among the methods based on traditional machine learning, Support Vector Machine (SVM) (Anh et al., 2012; Moshfeghi et al., 2013; Zulkifli et al., 2015), Bayesian network (Yoon and Chung, 2013), K-means algorithm (Mohanty and Swain, 2011), decision tree algorithm, K-Nearest Neighbor (KNN) (Heraz and Frasson, 2007) and other classifiers are exploited for emotion classification and recognition. However, traditional machine learning highly relies on manual feature extraction, which is time-consuming and labor-intensive. Because of the advantages of deep learning technology in automatic feature extraction, many EEG-based emotion recognition methods based on deep neural networks are proposed, such as Yang et al. (2018) using a hybrid neural network which combines 'Convolutional Neural Network (CNN)' and 'Recurrent Neural Network (RNN)' to effectively learn the compositional spatial-temporal representation of raw EEG streams, and Song et al. (2018) utilized the dynamical graph

convolutional neural network (DGCNN) to model multichannel EEG features and then classify the EEG emotion.

Since EEG signals are subject-specific, there are large variations between individuals (Jayaram et al., 2015). To solve the problem, Hwang et al. (2020) proposed an adversarial learning method for subject-independent EEG-based classification methods and achieved good performance on the SEED data set. Li et al. (2019a) proposed a multisource transfer learning method for cross-subject EEG emotion recognition, which can generalize existing models to a new person. But this method did not consider the invariant features of the domain, which would lead to the loss of part of the information. Therefore, many scholars introduce domain adaptation (DA) into deep learning models to obtain domain-invariant representations and achieve remarkable results. Zhu D. et al. (2019) provided a two-stage alignment framework for deep learning methods based on Multi-source Unsupervised Domain Adaptation (MUDA), which not only aligns the domain-specific distributions of each pair of source and target domains in multiple feature spaces to learn multiple domain-invariant representations but also align the domain-specific classifiers' output for target samples. This type of approach (Long et al., 2016) mainly matches the embedding of distributions by introducing an adaptive layer (Zhu J. et al., 2019). Another mainstream method based on deep learning is to directly utilize adversarial methods or generative adversarial methods, such as Wasserstein generative adversarial network domain adaptation (WGANDA) (Luo et al., 2018), few-label adversarial domain adaptation (FLADA) (Wang et al., 2021), joint distribution adaptation method (Li et al., 2019b), and so on. However, most of these methods are single-representation adaptation, which leads to the alignment of the source and target domains to concentrate on partial information (Zhu J. et al., 2019). Besides, due to the non-stationary between individual sessions (Sanei and Chambers, 2007), EEG-based emotion recognition still has the problem of cross-domain. Most of the existing studies aggregate the EEG data from different subjects and sessions into one source domain, ignoring the difference in the edge distribution of different EEG domains. Thus, how to achieve cross-subject and cross-session EEG-based emotion recognition is still a huge challenge.

To address these problems, we propose a cross-domain EEG emotion recognition method based on multi-source and multi-representation adaptation (MSMRA). In the data processing stage, the EEG signal is divided into multiple source domains to avoid destroying the edge distribution of the source domain data. Since the multi-source will complicate the network structure, we propose a two-partition method to aggregate two EEG data into one source domain. On this basis, the low-level domain-invariant features of different EEG data are extracted by the common feature extraction module. Then for multiple source domains, the multi-domain specific feature extractor (MDSFE) module enhances the representation capability of the source by aligning multiple representation distributions excavated from a hybrid structure. The multiple distributions are converted into a vector and fed into the classifier for final sentiment inference. Our extensive experiments on two public datasets show that



MSMRA can achieve remarkable performance compared with state-of-the-art competitors.

The contributions of this paper are summarized as follows.

- (1) We propose the MSMRA network to aggregate the EEG data of different subjects and sessions into multiple source domains, thereby improving the accuracy of emotion classification based on EEG in cross-domain scenarios.
- (2) A two-partition method is proposed to divide EEG data into multiple source domains, which not only significantly reduces the redundancy of the network structure but also expands the data of a single source to improve the emotion classification accuracy.
- (3) To obtain more high-level features of EEG signals, we constructed an MRA module to align the distribution of multiple different representations.
- (4) We conducted extensive experiments on two public datasets to validate the performance of MSMRA compared with state-of-the-art methods.

MATERIALS

BCI Process Flow for Emotion Recognition

As observed in **Figure 1**, the BCI process flow for emotion recognition usually consists of four-steps (Vasiljevic and Miranda, 2020; Gu et al., 2021). The subject is first stimulated with some stimuli, which are often movie clips containing various emotions, then the electrodes placed in the subject's skull are used to record the EEG signal. After data processing, the EEG data is finally fed into classification models.

Datasets

To evaluate the proposed methods, we conduct experiments on two public datasets, namely SEED and SEED-IV (Duan et al., 2013; Zheng and Lu, 2015; Zheng et al., 2019). In the SEED dataset, a total of 15 subjects (7 males and 8 females) participated in the experiment. Their EEG data was collected *via* 62 EEG electrodes when they were watching 15 Chinese movie clips

with negative, positive, and neutral emotions. To avoid subject fatigue, the entire experiment will not last for a long time, and the duration of each segment is about 4 min. Since the video can be understood without explanation and the video only triggers a single target emotion, all EEG signals will be classified into one of three emotional states (positive, neutral, and negative). The data collection lasts for 3 different periods at approximately 1-week intervals, corresponding to three sessions, so each subject has a total of 45 EEG data trials. In addition, to ensure that the collected EEG data and the movie clips presented to the subjects have the same emotional state, an additional subjective self-assessment will be performed on each subject after the subjects watch the movie. The relevant information about the SEED dataset is shown in **Table 1**.

Similar to SEED, SEED-IV contains a total of 15 subjects (7 males and 8 females) of 62-channel EEG data. These data are the four types of emotions collected by the subjects when they watch emotion-induced movie clips, namely, neutral, sad, fear, and happy. Each movie clip lasts about 2 min. A total of three sessions are collected. Each session includes 24 trials/movies for each subject. Compared to SEED that only uses EEG signals, the database also includes eye movement features recorded by SMI eye-tracking glasses. The relevant information about the SEED-IV dataset is shown in **Table 2**.

Experimental Scenarios

Due to the non-stationarity of the EEG signals between sessions and subjects (Sanei, 2013), the EEG signals of different subjects in the same session and the EEG signals of the same subject

TABLE 1 | Information about SEED dataset.

Attributes	Details Information
Source	BCMI laboratory
Sessions	Three
Subjects	Fifteen
Trials	Fifteen
Emotions	Positive Neutral Negative
Channels of recorded	62 EEG channel

TABLE 2 | Information about SEED-IV dataset.

Attributes	Details Information
Source	BCMI laboratory
Sessions	Three
Subjects	Fifteen
Trials	Twenty four
Emotions	Neutral Sad Fear Happy
Channels of recorded	62 EEG channel

in different sessions will have a certain degree of difference. Therefore, in the field of emotion recognition of EEG signals, domain adaptation can be divided into several situations:

Cross-Subject Transfer

In the same session, take the EEG data of the new last subject as the target domain, and the EEG data of all the remaining subjects as the source domain. In this article, since the SEED and SEED-IV datasets have a total of 3 sessions and 15 subjects, we will group the EEG data of the subjects in pairs to form a source field, there are 7 in total, and the final subject's EEG data is used as the target domain.

Cross-Session Transfer

In different session experiments of the same subject, we combine the EEG data of the previous two sessions into a source domain, and the EEG data of the last session is regarded as the target domain (see section "Methods" for details).

Electroencephalography Data Pre-processing

The collected data is first down-sampled to a sampling rate of 200 Hz, and then a band-pass frequency filter of 0–75 Hz is applied to filter noise and remove artifacts. After the EEG sequence is converted to the frequency domain, the differential entropy features of each frequency band (Delta, Theta, Alpha, Beta, and Gamma) and each channel (62 channels) are extracted (62 channel positions are shown in **Figure 2**), where the frequency of the Delta band is 1~4 Hz, and the frequency of the Theta band is 4–8 Hz, the frequency of the Alpha band is 8–14 Hz, the frequency of the Beta band is 14–31 Hz, and the frequency of the Gamma band is 31–50 Hz.

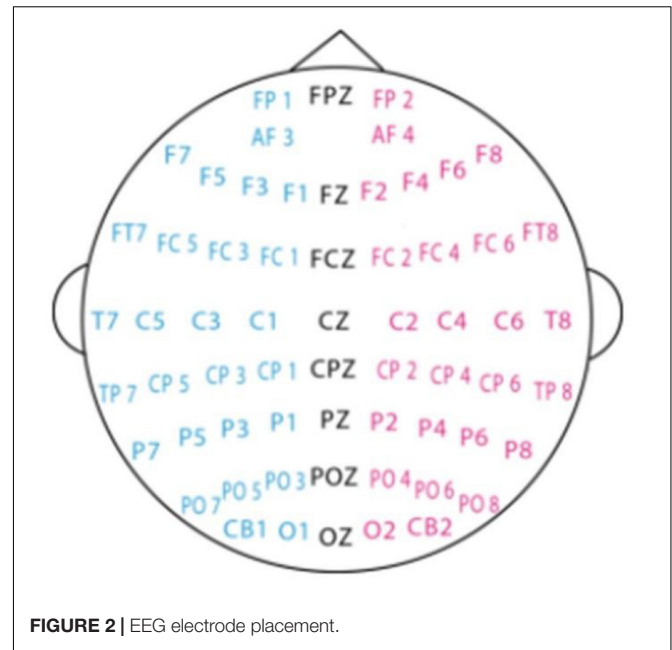
The differential entropy feature is defined as Equation 1:

$$DE = - \int_{-\infty}^{\infty} P(x) \ln [P(x)] dx \quad (1)$$

where DE represents the differential entropy features, x denotes a random variable, $P(x)$ indicates the probability density of x .

We assume that the EEG signal obeys a normal distribution $xN(\mu, \sigma^2)$, then the calculation of the DE feature can be simplified to Equation 2:

$$DE = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx \quad (2)$$

**FIGURE 2** | EEG electrode placement.

In this work, in both the Cross-Subject Transfer and Cross-Subject Transfer scenarios, we use DE features extracted from EEG raw data as input, and one data from one subject in one session for both databases is in the form of trials (SEED 15, SEED-IV 24) \times channel (62) \times band (5) \times samples, we then merge the channel and frequency band data into trials (SEED 15, SEED-IV 24) \times 310 \times samples. For the SEED dataset, 15 movie clips contain a total of 3,394 samples corresponding to each session, and for the SEED-IV dataset, 24 movie clips contain a total of 851/832/822 samples corresponding to three sessions.

METHODS

There are two significant concepts in the domain adaptation: the source domain represents a domain different from the test sample but has rich supervision information, and the target domain represents the domain where the test sample is located, without or only have a few labels (Pan and Qiang, 2010). The source domain and the target domain often belong to the same task, but the distribution is different. In this article, we regard the newly collected EEG data as the target domain, and all the remaining previously collected EEG data as the source domain. Given the data of the source domain and target domain, we are not simply aggregating all EEG data into a source domain, which ignores the internal differences of EEG data of different subjects and different sessions (Zhu D. et al., 2019), in contrast, we combine two groups of EEG data into multiple source domains, which not only considers the internal differences of EEG data but also expands the number of samples of source domain data, which can prove to be effective in improving the accuracy of sentiment classification based on EEG. Moreover, we believe that just projecting the source domain and target domain into a common feature subspace cannot cover all information

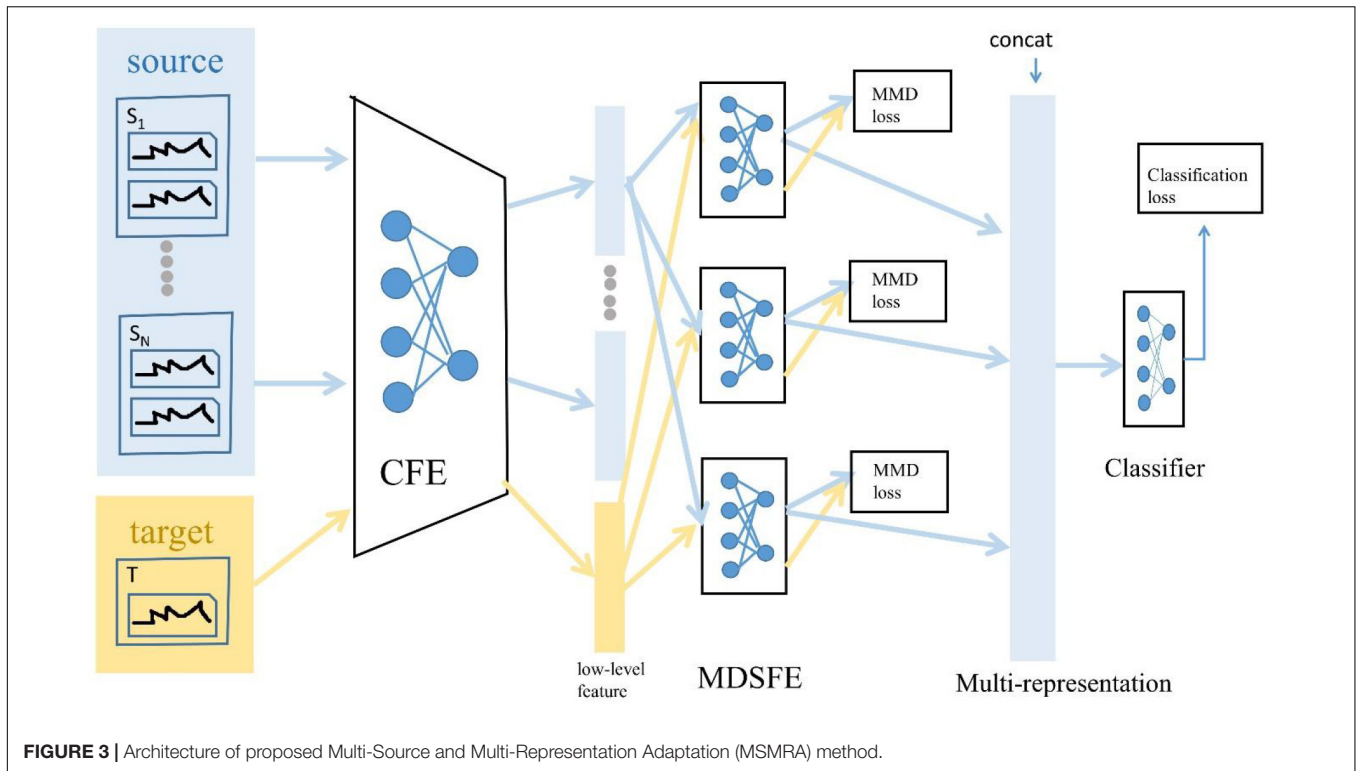


FIGURE 3 | Architecture of proposed Multi-Source and Multi-Representation Adaptation (MSMRA) method.

(Zhu J. et al., 2019), so we extract multiple representations from low-pixel images to obtain more information. The architecture is shown in Figure 3.

As shown in Figure 3, suppose we have N source domains, which can be defined as:

$$S = \{S_1, S_2, \dots, S_N\} \tag{3}$$

Where S denotes the collection of source domains, S_N means Nth source domain. Each source domain consists of two different EEG data, which is expressed as follows:

$$S_i = \{E_{2i-1}^S, E_{2i}^S\}_{i=1}^N \tag{4}$$

The target domain can be expressed as $T = \{E^T\}$. The data of the source domain and the target domain are fed into a common feature extractor module $c(\cdot)$ to extract low-level feature of both $\{F_i^S\}_{i=1}^N, F_i^T$:

$$F_i^S = c(S_i) \tag{5}$$

$$F_i^T = c(T) \tag{6}$$

Then the low-level features of the source domain and the target domain $\{F_i^S\}_{i=1}^N$ and F_i^T are fed into the multi-domain specific feature extractor module $\{h_j^i(\cdot)\}_{j=1}^r$ to obtain the multi-expression, respectively, $\{R_{ij}^S\}_{i=1}^N$ and R_{ij}^T .

$$R_{ij}^S = \{h_j^i(F_i^S)\}_{j=1}^r \tag{7}$$

$$R_{ij}^T = \{h_j^i(F_i^T)\}_{j=1}^r \tag{8}$$

The multi-domain-specific feature extractor has r different sub-domain-specific feature extractors, and each sub-domain-specific feature extractor can obtain various high-level feature expressions. Compared with a simple single representation, multiple representations can cover more information to better judge emotion classification. After obtaining the multiple representations of the source domain and the corresponding target domain $\{R_{ij}^S\}_{i=1}^N$ and R_{ij}^T , the maximum mean discrepancy (MMD) loss of the different expressions of the source domain and the corresponding target domain expression can be calculated. The MMD loss is a measure of the distance between the current source domain and the target domain (Borgwardt et al., 2006), and the final MMD loss is the sum of the MMD loss of each source domain expression and the corresponding target domain expression. The MMD loss can be formulated as Equation 9, and the final MMD loss can be defined as Equation 10.

$$MMD_j(R_{ij}^S, R_{ij}^T) = \left\| \frac{1}{n} \sum_{x=1}^n \phi(r_x^S) - \frac{1}{m} \sum_{y=1}^m \phi(r_y^T) \right\|_H^2 \tag{9}$$

$$MMD_{final}^j = \sum_{j=1}^r MMD_j(R_{ij}^S, R_{ij}^T) \tag{10}$$

Finally, the multi-representation vector $[R_{ij}^S]_{j=1}^r$ is connected into a new vector and fed into the classifier module $cls(\cdot)$ to realize the emotion classification of the EEG signal. Details of these modules are described below.

Source Domain Selection

Most existing domain adaptation methods simply aggregate all source domain data into one source domain, which often destroys the edge distribution of the source domain, or increases the difficulty of adaptation. Considering the practicality of the model, if the data of a single participant in a single session is simply divided into a source domain, this will greatly increase the complexity of the model and increase the number of parameters. Therefore, based on the above assumptions, we combine two different EEG data from one subject in one session to form a source domain, which not only satisfies the presumption that domain adaptation has a certain source marginal distribution, but also reduces the complexity of the model.

Common Feature Extractor

This module is used to map the original data of the source domain and the target domain to a common shared latent space, and then extract the low-level features of the two respectively, which are some low-level domain invariant features that represent similar characteristics of different EEG data.

Multi-Domain Specific Feature Extractor

After obtaining low-level features with domain invariance, we build N multi-domain specific feature extractors corresponding to N source domains. Each multi-domain specific feature extractor has r different domain-invariant feature extractors for learning multiple domain-invariant representations. Since diverse neural network structures can extract different representations from low-level features, we apply a multi-domain specific feature extractor to map each pair of source and target domain data to r latent spaces, and then in each latent space extract the high-level features of the two separately to obtain more domain-specific information. To apply DA and make this r pair of high-level features with different representations close to the corresponding latent space, we choose MMD to estimate the distance between the two domains in the latent space. The final MMD loss of the source domain is the MMD of all r latent spaces. Each source domain has r potential spaces, thus the final MMD loss is the sum of the MMD losses of all source domains and target domains, which can be expressed by Equation 11.

$$MMD_{final} = \sum_{i=1}^N MMD_{final}^i \quad (11)$$

Classifier

The multiple high-level feature representations obtained from the multi-domain specific feature extractor are connected into a new vector and sent to the classifier for classification. Each source domain corresponds to a softmax classifier. We use cross-entropy to evaluate classification loss, as shown in Equation 12:

$$L_{cls}^i = - \sum_{a=1}^M c_a * \log(p_a) \quad (12)$$

In summary, we first synthesize N source domains into two sets of $2N$ data from one subject in one session and consider the

last collected single-session single-subject EEG data as the target domain, and then all source domains and target domain data are fed into the general feature extractor to extract low-level features. Next, we use a multi-domain specific feature extractor to map the low-level features of the source and target domains to multiple different latent spaces and apply MMD loss to measure the high-level feature representations of the source and target domains in various latent spaces. Finally, we connect different expressions into a new vector and send it to the classifier for classification. The final loss function can be expressed as Equation 13:

$$L_{loss}^i = L_{cls}^i + \alpha MMD_{final}^i \quad (13)$$

The total loss function of all source and target domain data can be defined as Equation 14:

$$L_{loss} = \sum_{i=1}^N L_{loss}^i \quad (14)$$

The training is based on Equation 14 and follows the algorithm shown in **Algorithm 1**. Minimizing this formula is to minimize the classification loss and MMD loss so that the distance between the source domain and the target domain can be as small as possible in different potential spaces, and the classification prediction is as close as possible to the actual label.

EXPERIMENTS

In our experiment, we use the public SEED dataset and SEED-IV dataset to classify emotions. In addition, we not only compare our proposed method with the baseline method but also conduct a large number of ablation experiments and exploratory experiments.

ALGORITHM 1 Overview of MSMRA module.

Input:

S, T : the source domain and the target domain of EEG data
epoch: the number of epoch

for $e \leftarrow 1$:epoch **do**

 Draw one batch of training samples from each source domain S_i and the target domain T ;

 Obtain low-level features of the source and target domains;

$F_i^s = c(S_i), F_i^t = c(T)$

 Obtain multi-Representations of low-level features in the source and target domains;

$R_{ij}^s = \{h_j^i(F_i^s)\}_{j=1}^r, R_{ij}^t = \{h_j^i(F_i^t)\}_{j=1}^r$

 Calculate the final MMD loss;

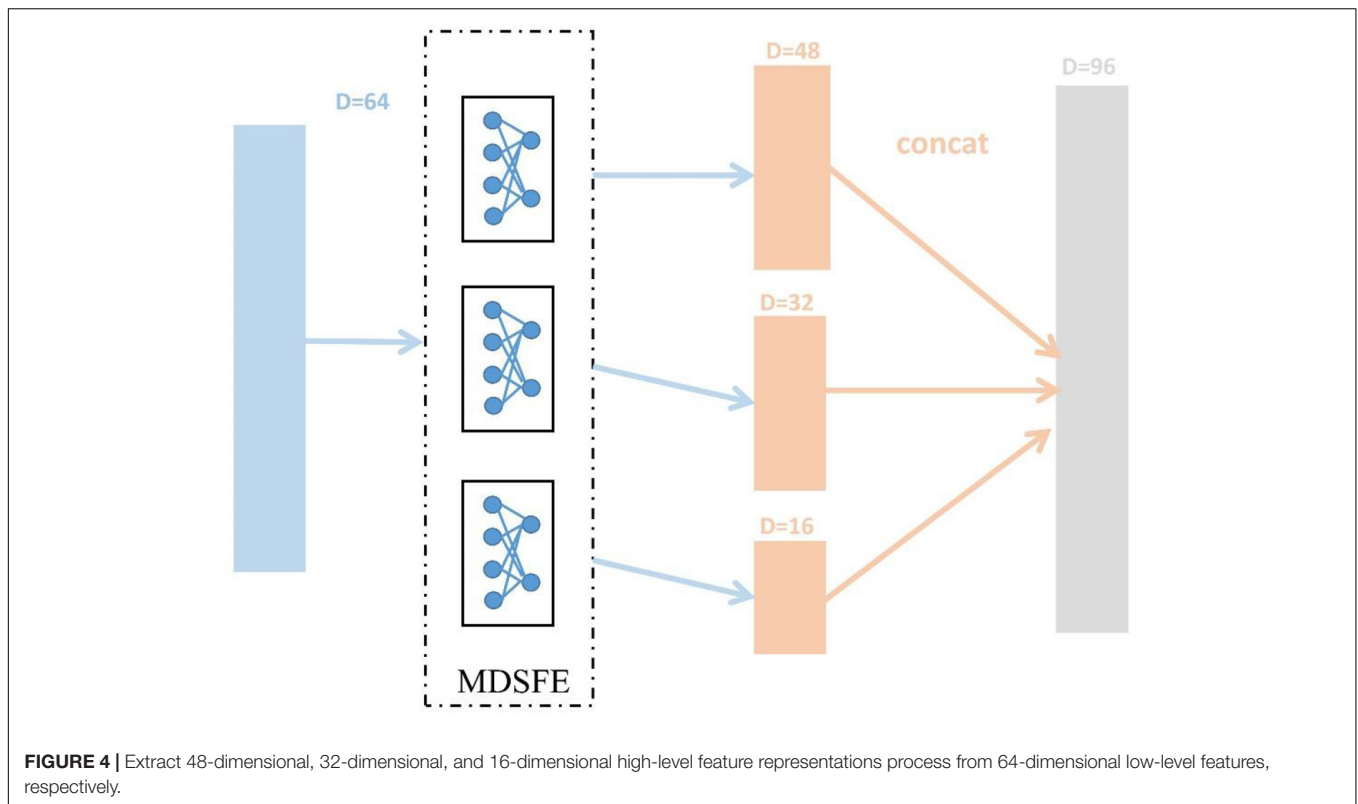
$MMD_j(R_{ij}^s, R_{ij}^t) = \|\frac{1}{n} \sum_{x=1}^n \phi(r_x^s) - \frac{1}{m} \sum_{y=1}^m \phi(r_y^t)\|_H^2$
 $MMD_{final}^i = \sum_{j=1}^r MMD_j(R_{ij}^s, R_{ij}^t)$

 Concat the multi-representations high-level features and send them to the classifier to calculate the cross-entropy loss L_{cls}^i

 The total loss is $L_{loss}^i = L_{cls}^i + \alpha MMD_{final}^i$

 Update model by minimizing the total loss

end for



Implementation Details

The SEED dataset and SEED-IV dataset are first pre-processed (mentioned in section “EEG Data Pre-processing”) to obtain DE features, and reshaped into trials $(\text{SEED } 15, \text{ SEED-IV } 24) \times 310 \times \text{samples}$. Then we normalize the dimensions of samples, and the ablation experiment in section “Ablation Study” will confirm the effect. After that, we divide the data into N source domains and one target domain in pairs in the manner mentioned in section “Methods.” As described in section “Methods,” the various modules of MSMRA have many details. For the common feature extraction module,

since we do not have as many samples as the original data, so in this layer, we take a simple three fully connected layers, which reduce the dimension from 310 (62×5) to 64, each fully connected layer is followed by a LeakyReLU activation function (Xu et al., 2015). We also evaluated the ReLU activation function (Nair and Hinton, 2010). Since EEG data is very sensitive, using ReLU will delete many values less than 0, which often loses a lot of information, so we choose LeakyReLU as the activation function. For the multi-domain specific feature extractor module, we use 3 fully connected layers to downsample the low-level feature vector from 64-dimensional to 48-dimensional, 32-dimensional, and 16-dimensional high-level feature representations and then connect them into a vector, as shown in **Figure 4**. Like the common feature extraction module, 3 fully connected layers are followed by a LeakyReLU activation function. The vector connected by the high-level features is fed into the final classifier, which is a simple fully connected layer, reducing the number of emotion categories (SEED 3, SEED-IV 4) corresponding to the dataset from 96 dimensions. During the network training process, the initial value of the learning rate is 0.01, the batch size is 256, and the epoch is set to 50. The Adam optimizer is used for gradient descent (Kingma and Ba, 2014). In addition, we dynamically adjust the coefficient to achieve the effect of giving priority to the classification results, and the coefficient can be expressed as Equation 15.

TABLE 3 | Comparisons of the average accuracies and standard deviations of cross-session and cross-subject scenarios on SEED database among the various methods.

Dataset	Method	Cross-session	Cross-subject
SEED	DGCNN	–	79.95 ± 9.02
	DDC	81.53 ± 6.83	68.99 ± 3.23
	DAN	79.93 ± 7.06	65.84 ± 2.25
	DCORAL	76.86 ± 7.61	66.29 ± 4.53
	DANN	–	79.19 ± 13.14
	PPDA	–	86.70 ± 7.10
	MS-MDA	88.56 ± 7.80	89.63 ± 6.79
	MSMRA (Ours)	90.30 ± 5.26	87.62 ± 7.53

Bold indicates the maximum average accuracies of cross-session and cross-subject scenarios among the various methods.

$$\alpha = \frac{2}{1 + e^{-10 \cdot i / \text{epoch}}} - 1 \quad (15)$$

TABLE 4 | Comparisons of the average accuracies and standard deviations of cross-session and cross-subject scenarios on SEED-IV database among the various methods.

Dataset	Method	Cross-session	Cross-subject
SEED-IV	DDC	57.63 ± 11.28	37.71 ± 6.36
	DAN	55.14 ± 12.79	32.44 ± 9.02
	DCORAL	44.63 ± 11.38	37.43 ± 3.08
	MS-MDA	61.43 ± 15.71	59.34 ± 5.48
	MSMRA (Ours)	72.38 ± 10.12	69.77 ± 7.37

Bold indicates the maximum average accuracies of cross-session and cross-subject scenarios among the various methods.

TABLE 5 | Ablation study of MSMRA on SEED and SEED-IV.

Dataset	Method	Cross-session	Cross-subject
SEED	Ours full	90.30 ± 5.26	87.62 ± 7.53
	w/o normalization	80.21 ± 9.95	84.12 ± 6.17
	w/o MDSFE	89.55 ± 5.17	87.17 ± 5.41
	w/o normalization + MDSFE	77.02 ± 11.11	80.88 ± 7.22
SEED-IV	Ours full	72.38 ± 10.12	69.77 ± 7.37
	w/o normalization	43.39 ± 6.97	52.27 ± 5.18
	w/o MDSFE	71.97 ± 12.53	60.19 ± 9.60
	w/o normalization + MDSFE	43.33 ± 4.79	50.94 ± 3.24

Bold indicates the maximum average accuracies of cross-session and cross-subject scenarios among the various methods.

BaseLines

We compare the MSMRA model with the latest various competitors on the SEED dataset and SEED-IV dataset, including Deep domain confusion (DDC) (Tzeng et al., 2014), which treats the MMD algorithm as an adaptive metric, domain adaptation model (DAN) (Li et al., 2018), which adopts deep adaptation network to EEG-based emotion recognition, Deep coral (DCORAL) (Sun and Saenko, 2016), which aligns the second-order statistical features of the source domain and target domain distribution by linear transformation method, Domain-Adversarial Training of Neural Networks (DANN) (Ganin et al., 2017), which introduces the idea of adversarial network to the field of transfer learning, Plug-and-play domain adaptation (PPDA) (Zhao et al., 2021), which proposes a plug-and-play domain adaptation method for reducing the calibration time. Dynamical Graph Convolutional Neural Networks (DGCNN) (Song et al., 2018) and Multisource Marginal Distribution Adaptation (MS-MDA) (Chen et al., 2021) which applies a multi-source domain method.

Results

All the experimental results of the two data sets are shown in **Tables 3, 4**, respectively. In the experiment, except for some results directly taken from the original paper (MS-MDA, DGCNN), all the hyperparameters are the same. From these results, we have the following profound observations.

- In the SEED data set, our proposed method MSMRA surpasses most of the current state-of-the-art methods in Cross-subject and Cross-session scenarios. It can be seen

that our method improves the accuracy of at least close to 2% compared to other methods in the Cross-session scene, and for the first time exceeded the 90% accuracy rate, as far as we know. In the Cross-subject scenario, although our method is 2% lower than MS-MDA, it is better than all other methods, which shows that our method is competitive.

- On the SEED-IV dataset, it can be clearly seen that our proposed method significantly exceeds all other competitors, and each has improved by at least 11 and 10% compared to others in Cross-session and Cross-subject scenarios, it is a piece of exciting news. From the results, we can infer that our method can also achieve good classification results on small sample datasets because multiple high-level features containing rich information are extracted from the MDSFE module.

Ablation Study

To further understand the MSMRA model we proposed, we performed the ablation experiments of removing the initial normalization of the data and removing the MDSFE module on the SEED and SEED-IV datasets respectively, and evaluate the performance of the ablated model. The hyperparameters of all ablation experiments are consistent, and the experimental results are shown in **Table 5**. As shown in the table, the first row of the two datasets is the complete method we proposed, and we can see that the accuracy rate is the highest; the second row removes the initial data normalization, which makes the model prone to be affected by outliers resulting in a decrease in accuracy; the third row removes the multi-domain specific feature extraction module, which weakens the feature representation ability of the model. In the fourth row, the initial data normalization module and the multi-domain specific feature extraction module are removed. It can be seen that the experimental accuracy is lower than the removal of a single ablation module. Moreover, notice that although a certain module is removed, our proposed method is still competitive with other methods.

CONCLUSION

In general, the current emotion classification based on EEG signals is hindered due to the instability of the internal data of Cross-session and Cross-subject when collecting data. Domain adaptation increases the classification accuracy of cross-domain data by finding a mapping relationship to map the source domain and target domain to a latent space, and minimizing the distance between the two. Based on the above, we propose a multi-source and multi-representation domain adaptive network (MSMRA), which regards EEG signals as multiple source domains rather than simply as one source domain, which satisfies the assumption that the source domain data has a certain marginal distribution. It can be seen from **Tables 3, 4** that the classification accuracy of our proposed method significantly exceeds that of the latest single-source domain method. Since the multi-source domain method is easy to cause the network to be bloated and complicated, we propose a pairwise combination method to solve this problem. Moreover, considering that the representation

distribution extracted in a single structure of aligning may only contain partial information, we propose a multi-domain specific feature extraction module, which can extract multiple high-level features of different dimensions. **Table 5** shows that this module is helpful to improve the accuracy of emotion classification. Future work will focus on how to divide multi-source domains and how to extract more effective multi-domain specific features.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://bcmi.sjtu.edu.cn/~seed/index.html>.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s)

REFERENCES

- Anh, V. H., Van, M. N., Ha, B. B., and Quyet, T. H. (2012). "A real-time model based Support Vector Machine for emotion recognition through EEG," in *International Conference on Control, Automation and Information Sciences (ICCAIS)*, (Saigon: IEEE), 191–196. doi: 10.1109/ICCAIS.2012.6466585
- Borgwardt, K. M., Gretton, A., Rasch, M. J., Kriegel, H. P., Schölkopf, B., and Smola, A. A. J. (2006). Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics* 22, e49–e57. doi: 10.1093/bioinformatics/btl242
- Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *Int. J. Hum. Comput. Stud.* 59, 213–225. doi: 10.1016/S1071-5819(03)00050-8
- Chen, H., Jin, M., Li, Z., Fan, C., Li, J., and He, H. (2021). MS-MDA: multisource Marginal Distribution Adaptation for Cross-subject and Cross-session EEG Emotion Recognition. *Front. Neurosci.* 15:778488. doi: 10.3389/fnins.2021.778488
- Duan, R. N., Zhu, J. Y., and Lu, B. L. (2013). "Differential entropy feature for EEG-based emotion classification," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on IEEE*, (Piscataway: IEEE), 81–84.
- Durupinar, F., Kapadia, M., Deutsch, S., Neff, M., and Badler, N. I. (2016). Perform: perceptual approach for adding ocean personality to human motion using laban movement analysis. *ACM Trans. Graph.* 36, 1–16. doi: 10.1145/2983620
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al. (2017). Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17, 2096–2030. doi: 10.1007/978-3-319-58347-1_10
- Gu, X., Cao, Z., Jolfaei, A., Xu, P., and Lin, C. T. (2021). EEG-based brain-computer interfaces (bcis): a survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1645–1666. doi: 10.1109/TCBB.2021.3052811
- Heraz, A., and Frasson, C. (2007). Predicting the three major dimensions of the learner's emotions from brainwaves. *international journal of computer science. Int. J. Comput. Sci.* 2, 187–193.
- Hwang, S., Ki, M., Hong, K., and Byun, H. (2020). "Subject-Independent EEG-based Emotion Recognition using Adversarial Learning," in *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, (Korea: IEEE), 1–4. doi: 10.1109/BCI48061.2020.9061624
- Jayaram, V., Alamgir, M., Altun, Y., Schölkopf, B., and Grosse-Wentrup, M. (2015). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545
- Kingma, D., and Ba, J. (2014). "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, Ithaca, NY.
- Li, H., Jin, Y. M., Zheng, W. L., and Lu, B. L. (2018). "Cross-Subject Emotion Recognition Using Deep Adaptation Networks," in *International Conference on Neural Information Processing*, (Cham: Springer), 403–413. doi: 10.1007/978-3-030-04221-9_36
- Li, J., Qiu, S., Du, C., Wang, Y., and He, H. (2019a). Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Trans. Cogn. Dev. Syst.* 12, 344–353. doi: 10.1109/TCDS.2019.2949306
- Li, J., Qiu, S., Shen, Y., Liu, C., and He, H. (2019b). Multisource transfer learning for cross-subject EEG emotion recognition. *IEEE Trans. Cybern.* 50, 3181–3293. doi: 10.1109/TCYB.2019.2904052
- Long, M., Zhu, H., Wang, J., and Jordan, M. I. (2016). "Deep transfer learning with joint adaptation networks," in *Proceedings of the 34th International Conference on Machine Learning, International Conference on Machine Learning (ICML) 2017*, (Sydney: ICML).
- Luo, Y., Zhang, S. Y., Zheng, W. L., and Lu, B. L. (2018). WGAN domain adaptation for EEG-based emotion recognition. *Int. Conf. Neural Inf. Process.* 11305, 275–286. doi: 10.1007/978-3-030-04221-9_25
- Mohanty, S., and Swain, B. K. (2011). Emotion Recognition using Fuzzy K-Means from Oriya Speech. *Int. J. Comput. Commun. Technol.* 2, 188–192. doi: 10.47893/IJCCT.2011.1066
- Moshfeghi, M., Bartaula, J. P., and Bedasso, A. T. (2013). *Emotion Recognition from EEG Signals using Machine Learning*. Available online at: <http://urn.kb.se/resolve?urn=urn:nbn:se:bth-4147>
- Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, (Haifa: ICML).
- Niemic, C. P. (2004). Studies of emotion: a theoretical and empirical review of psychophysiological studies of emotion. *J. Undergrad. Res.* 1, 15–18.
- Pan, S. J., and Qiang, Y. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Said, Y., and Barr, M. (2021). Human emotion recognition based on facial expressions via deep learning on high-resolution images.

for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JC wrote the main manuscript and conducted the experiments. XH and CY participated in the writing of the manuscript and modified the English grammar of the article. SC and ZL made the experiments. ZW analyzed the results. All authors reviewed the manuscript.

FUNDING

This work has been supported by a grant from Natural Science Foundation of Fujian Province (2020J02063), Xiamen Science and Technology Bureau Foundation of Science and Technology Project for Medical and Healthy (3502Z20209005), and Xiamen Science and Technology Bureau Foundation of Science and Technology Project for Medical and Healthy (3502Z20214ZD1013).

- Multimed. Tools Appl.* 80, 25241–25253. doi: 10.1007/s11042-021-10918-9
- Sanei, S. (2013). *EEG Signal Processing*. Springer: Netherlands.
- Sanei, S., and Chambers, J. A. (2007). EEG signal processing. *Comput. Intell. Neurosci.* 2, 1178–1181. doi: 10.1002/9780470511923
- Senecal, S., Louis, C., Andreas, A., and Nadia, M. (2016). Continuous body emotion recognition system during theater performances. *Comput. Animat. Virtual Worlds* 27, 311–320. doi: 10.1002/cav.1714
- Song, T., Zheng, W., Song, P., and Cui, Z. (2018). “EEG emotion recognition using dynamical graph convolutional neural networks,” in *IEEE Transactions on Affective Computing*, (Piscataway: IEEE). doi: 10.1109/TAFFC.2018.2817622
- Sun, B., and Saenko, K. (2016). *Deep CORAL: Correlation Alignment for Deep Domain Adaptation*. Cham: Springer International Publishing, 443–450.
- Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., and Zafeiriou, S. (2016). “Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Shanghai: IEEE), 5200–5204. doi: 10.1109/ICASSP.2016.7472669
- Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., and Darrell, T. (2014). Deep domain confusion: maximizing for domain invariance. *arXiv [Preprint]*. arXiv:1412.3474
- Vasiljevic, G., and Miranda, D. (2020). Brain-computer interface game based on consumer-grade EEG devices: a systematic literature review. *Int. J. Hum. Comput. Interact.* 36, 105–142. doi: 10.1080/10447318.2019.1612213
- Wang, Y., Liu, J., Ruan, Q., Wang, S., and Wang, C. (2021). Cross-subject EEG emotion classification based on few-label adversarial domain adaption. *Expert Syst. Appl.* 185:115581. doi: 10.1016/j.eswa.2021.115581
- Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv [Preprint]*. arXiv:1505.00853
- Yang, Y., Wu, Q., Ming, Q., Wang, Y., and Chen, X. (2018). “Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network,” in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, (Piscataway: IEEE), 1–7. doi: 10.1109/IJCNN.2018.8489331
- Yoon, H., and Chung, S. (2013). EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Comput. Biol. Med.* 43, 2230–2237. doi: 10.1016/j.compbiomed.2013.10.017
- Zhao, L. M., Yan, X., and Lu, B. L. (2021). Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition. *Proc. AAAI Conf. Artif. Intell.* 35, 863–870.
- Zheng, W.-L., Liu, W., Lu, Y., Lu, B. L., and Cichocki, A. (2019). Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176
- Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497
- Zhu, Y., Zhuang, F., and Wang, D. (2019). Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. *Proc. AAAI Conf. Artif. Intell.* 33, 5989–5996. doi: 10.1609/aaai.v33i01.33015989
- Zhu, Y., Zhuang, F., Wang, J., Chen, J., and He, Q. (2019). Multi-representation adaptation network for cross-domain image classification. *Neural Netw.* 119, 214–221. doi: 10.1016/j.neunet.2019.07.010
- Zulkifli, N., Ali, S., Ahmad, S. A., and Islam, M. S. (2015). Review on Support Vector Machine (SVM) classifier for human emotion pattern recognition from EEG signals. *Asian J. Inf. Technol.* 14, 135–146. doi: 10.3923/ajit.2015.135.146

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

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cao, He, Yang, Chen, Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.