



## OPEN ACCESS

## EDITED BY

Vassiliy Tsytarev,  
University of Maryland, United States

## REVIEWED BY

Hatim Aboalsamh,  
King Saud University, Saudi Arabia  
Lei Zhu,  
Hangzhou Dianzi University, China  
Lei Su,  
Kunming University of Science and  
Technology, China

## \*CORRESPONDENCE

Wei Zhong  
✉ wzhong@cuc.edu.cn

RECEIVED 14 September 2023

ACCEPTED 26 January 2024

PUBLISHED 14 February 2024

## CITATION

Gong M, Zhong W, Ye L and Zhang Q (2024)  
MISNet: multi-source information-shared EEG  
emotion recognition network with  
two-stream structure.  
*Front. Neurosci.* 18:1293962.  
doi: 10.3389/fnins.2024.1293962

## COPYRIGHT

© 2024 Gong, Zhong, Ye and Zhang. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# MISNet: multi-source information-shared EEG emotion recognition network with two-stream structure

Ming Gong<sup>1</sup>, Wei Zhong<sup>2\*</sup>, Long Ye<sup>2</sup> and Qin Zhang<sup>2</sup>

<sup>1</sup>Key Laboratory of Media Audio and Video (Communication University of China), Ministry of Education, Beijing, China, <sup>2</sup>State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China

**Introduction:** When constructing machine learning and deep neural networks, the domain shift problem on different subjects complicates the subject independent electroencephalography (EEG) emotion recognition. Most of the existing domain adaptation methods either treat all source domains as equivalent or train source-specific learners directly, misleading the network to acquire unreasonable transfer knowledge and thus resulting in negative transfer.

**Methods:** This paper incorporates the individual difference and group commonality of distinct domains and proposes a multi-source information-shared network (MISNet) to enhance the performance of subject independent EEG emotion recognition models. The network stability is enhanced by employing a two-stream training structure with loop iteration strategy to alleviate outlier sources confusing the model. Additionally, we design two auxiliary loss functions for aligning the marginal distributions of domain-specific and domain shared features, and then optimize the convergence process by constraining gradient penalty on these auxiliary loss functions. Furthermore, the pre-training strategy is also proposed to ensure that the initial mapping of shared encoder contains sufficient emotional information.

**Results:** We evaluate the proposed MISNet to ascertain the impact of several hyper-parameters on the domain adaptation capability of network. The ablation experiments are conducted on two publically accessible datasets SEED and SEED-IV to assess the effectiveness of each loss function.

**Discussion:** The experimental results demonstrate that by disentangling private and shared emotional characteristics from differential entropy features of EEG signals, the proposed MISNet can gain robust subject independent performance and strong domain adaptability.

## KEYWORDS

EEG signals, emotion recognition, transfer learning, multi-source domain, domain adaptation

## 1 Introduction

Emotion is critical in influencing people's decision-making, social interaction and evaluation of things (Dolan, 2002). By incorporating emotional analysis into human-machine interactions, the machines can better understand humanity and become more natural (Picard, 2001). Numerous studies have been conducted on emotion recognition based on various modes, such as facial expressions (Ko, 2018), speech (Schuller, 2018) and electrophysiological signals.

Electroencephalography (EEG) stands out among these signals due to its objective properties and high temporal resolution benefits (Yang et al., 2021). Specifically, EEG-based affective brain-computer-interfaces (aBCIs) (Mühl et al., 2014) aim to detect affective states from EEG signals and use them in various applications, such as estimating driver drowsiness to improve driving safety (Wu et al., 2016; Cui et al., 2019; Jiang et al., 2020) and establishing an objective detection system for depression (Cai et al., 2020) or post-traumatic stress disorder (Rozgic et al., 2014) to enable self-diagnosis.

Regarding EEG emotion recognition, depending on the head size, body state and experimental environment, the structural and functional variability of the brain may vary between the subjects, resulting in substantial differences in the collected EEG signals (Samek et al., 2013). Traditional machine-learning algorithms usually train a classifier such as support vector machines (Zheng and Lu, 2016) or random forests (Gupta et al., 2018), by utilizing data from a limited number of objects. Nevertheless, due to the EEG signals do not satisfy the independent and identically distributed condition which is caused by the individual difference and non-stationary properties, directly using the subject-dependent models to detect the emotional states of a new subject decreases recognition accuracy. Although collecting a large amount of labeled data from the new subject and using them to fine-tune the classifier is an obvious solution, it is time-consuming and degrades the subjects' experience significantly (Zhao et al., 2021). Hence, this strategy cannot be utilized in the practical aBCI applications.

The unsupervised domain adaptation is an alternative method to align different distribution domains, bridging the existing labeled subjects and new unlabeled ones by identifying their similarities (Wang and Chen, 2021). However, without access to the target domain, it is challenging to train a well-generalized network (Blanchard et al., 2011; Zhou et al., 2023). In contrast, the performance of unsupervised domain adaptation approaches is typically enhanced in the training phase by using unlabeled data from the target domain and employing instance-based, model-based, or feature-based (Wang and Chen, 2021) methods.

Compared with traditional machine learning, using deep learning to solve domain adaptation problems has relatively low requirements for trainers to select features. Based on the significant advancements in computer vision, speech recognition and natural language processing, we believe that the deep learning methods have potential in EEG emotion recognition. Regarding EEG emotion recognition, there have been sufficient studies on subject-dependent experiments (Kim and André, 2008; Ding et al., 2020; Nath et al., 2020; Pan et al., 2021; Zhang et al., 2022; Song et al., 2023). Several experimental results indicate that deep learning has a great potential for solving domain adaptation problems (Craik et al., 2019; Roy et al., 2019). When using deep learning in aBCI domain adaptation applications, most works regarded all source domains as being the same (Li H. et al., 2018; Li Y. et al., 2018; Luo et al., 2018). Hence all source domains should be merged into the common domain to extract features. This strategy disregards the distribution difference inside the source domains, resulting in the model being unable to train to the optimal effect. When there are outlier source domains, the model is difficult to converge, leading to "negative transfer". On the other hand, some researchers identified

the distribution difference mentioned above and trained domain-specific networks directly (Chen et al., 2021a,b; Luo and Lu, 2021), marginally improving recognition performance but overlooking the commonality among source domains. Furthermore, most of these approaches need to judge the distance between the features of target domain and those of each source domain to select one or several similar source domains, and weight the predictions to form the final prediction. When one source domain has larger distance to others, which means there is an outlier source domain, it may occur that one private domain mapping of target domain is far from other private domain mappings, and the model performance would decrease. Therefore, it is necessary to consider the individual difference and group commonality among the multi-source domains, further improving recognition performance.

This paper considers the individual difference and group commonality of multi-source domains and proposes the multi-source information-shared EEG emotion recognition network based on marginal distribution. In the proposed network, the domain-specific and domain-shared features are extracted and combined dynamically to alleviate the negative transfer problem. Specifically, we first integrate a pre-training strategy into the network to maximize the utilization of current source domain data and reasonably initialize the network, further enhancing its stability. Then, we extract domain-specific features by using private encoders and domain-shared features by a pre-trained shared encoder to represent the individuality and commonality of EEG signals from different domains. Besides employing the maximum mean discrepancy to align the marginal distributions between the source and target domains, two auxiliary loss functions are also designed to improve the astringency of network and align the distributions of private target domains. These loss functions can further enhance the mapping capability of private encoders by considering the information of other private domains. Moreover, rather than heuristically altering the weights of classifiers, we integrate the outputs of classifiers according to the domain-specific and domain-shared feature distributions, thereby dynamically optimizing the network. The experimental results on the SEED and SEED-IV datasets validate the performance of proposed method.

The main contributions of this paper are summarized as follows:

- We propose an efficient EEG emotional recognition network that incorporates the individual difference and group commonality of multi-source domains.
- We design a two-stream training structure and loop iteration strategy to compute two auxiliary loss functions  $L_{was-gp}$  and  $L_{diff-gp}$  for aligning the marginal distributions of domain-specific and domain-shared features in target domains. Furthermore, the gradient penalty is also constrained on the above two losses to improve the stability of network.
- We introduce the subject-dependent pre-training process to initialize the shared encoder with reasonable parameters, which supplies emotional information to the shared domain.

The remainder of this paper is organized as follows. Section 2 introduces the related works on domain

adaptation and EEG-based subject independent emotion recognition. Section 3 proposes the multi-source information-shared network and illustrates the corresponding training process. Section 4 presents the experimental settings and implementation details. Subsequently in Section 5, the results of the ablation experiments are analyzed and the comparisons are made on the SEED and SEED-IV datasets. Finally, Section 6 concludes this work and suggests future research directions.

## 2 Related work

This section briefly reviews the concept and methods of domain adaptation and then introduces the relevant work on EEG-based subject independent emotion recognition.

### 2.1 Domain adaptation

In domain adaptation, which is a rapidly growing transfer learning direction, the labeled source and unlabeled target domains share the same features and categories. The domain adaptation focuses on using the source domain knowledge to process the target domain features when the source and target domain distributions are different (Wang and Chen, 2021). Adopting deep learning for domain adaptation can automatically extract more expressive features and meet the end-to-end needs of practical applications. Typically, three categories are available: instance-based learning, model-based learning, and feature-based learning. The instance-based learning aims to select and weight samples from the source and target domains (Blitzer et al., 2006; Li et al., 2016). The objective of model-based learning is to transfer parameters between different models. By mapping the different probability distributions of the source and target domains, the feature-based learning characterizes the similarity between the source and target domains, which can be classified as marginal, conditional, joint or dynamic distribution adaptation.

In many practical applications involving multiple source domains, the multi-source domain adaptation methods can be used to transfer knowledge from multiple domains and consider domain shifts among source domains to achieve better transfer results. Recently, Zhao et al. (2018) bridged deep learning and multi-source domain adaptation by developing a multiple-domain discriminator to align the features of source and target domains, which is a typical adversarial discriminative method. Xu et al. (2018) constructed multiple domain discriminators and classifiers for each source-target domain pair. Then the target labels are voted according to the distribution-weight combining rule. Zhu et al. (2019) extracted distinct source domains into distinct feature spaces and aligned the source and target domains across each feature space. Moreover, they reduced the variance of the classifier output through consistency regularization to directly average the output of classifier and avoid the artificial setting.

### 2.2 EEG-based subject independent emotion recognition

Since the differences in gender, body state and experimental environment between individuals will lead to different neurophysiological activity patterns, the EEG signals of different subjects do not satisfy the independent and identically distributed condition. In this scenario, the issue of domain shift has arisen, that is, under the same emotional stimulus, different individuals may have different EEG responses, resulting in inconsistent distribution of collected EEG signals. The domain shift problem is the main challenge that the subject-independent algorithms need to address. It not only appears in the different sources of EEG data, but also may appear in the same EEG source due to psychological changes of the participants or technical factors, which greatly limiting the performance of the model.

To reduce inter-subject variability, transfer learning in EEG emotion recognition has two primary branches: domain adaptation and domain generalization. Through the data manipulation, representation learning and learning strategy, the domain generalization aims to learn a model from multiple source domains that generalizes to unseen target domains (Wang et al., 2023). Since the domain generalization methods do not utilize the information of target domain during training, they rarely obtain high recognition accuracy. In contrast, the domain adaptation methods use the information from target domain to transfer knowledge while minimizing domain shifts between the source and target domains. Zheng and Lu (2016) applied transfer component analysis (TCA) and transductive parameter transfer (TPT) to the subject independent EEG emotion recognition on the SEED dataset. Li H. et al. (2018) suggested an alternative method by employing the domain-adversarial neural network (DANN), which involves the adversarial training of feature encoder and domain classifier. Luo et al. (2018) proposed the wasserstein GAN domain adaptation network (WGANDA) by using the gradient penalty to alleviate the domain shift problem. By considering multi-source domain adaptation, Luo and Lu (2021) proposed the wasserstein-distance-based multi-source adversarial domain adaptation (wMADA), which regarded different subjects as different domains and designed an adaptive weight strategy considering the relationship between each domain. Zhao et al. (2021) developed a plug-and-play domain adaptation (PPDA) network, which disentangles the emotional information by considering the domain-specific and domain-invariant information simultaneously. Chen et al. (2021b) took the source data with different marginal distributions into account and proposed a multi-source EEG-based emotion recognition network (MEERNet). Later, they used the disc-loss to improve domain adaptation ability and proposed another multi-source marginal distribution adaptation (MS-MDA) network for subject independent and cross-session EEG emotion recognition (Chen et al., 2021a).

It should be noticed that, most of the existing domain adaptation methods mentioned above either treat all source domains as equivalent or train source-specific learners directly, misleading the network to acquire unreasonable transfer knowledge and thus resulting in negative transfer. Therefore,

this paper considers the feature individuality and commonality of distinct domains and weights similar domains based on their feature distributions, enhancing recognition performance.

### 3 Methods

In this section, we present the entire architecture and its data transmission process. And then the involved modules are analyzed in detail, and the loss functions are also designed by aligning different domains.

#### 3.1 Framework

The main challenge we aim to address is the domain shift problem caused by the non-stationary of EEG signals and the individual differences among users. [Figure 1](#) shows the framework of the proposed multi-source information-shared network (MISNet) based on marginal distribution, which comprises five components: common encoder, private encoders, shared encoder, private classifiers and shared classifier. In [Figure 1](#), the pink lines, shapes and arrows represent the path of the EEG data matrix  $\mathbf{X}_S$  from the source domains, and the green lines, shapes and arrows represent that of the EEG data matrix  $\mathbf{X}_T$  from the target domain. For each domain, the low-level features are firstly extracted by the common encoder, and then the private encoders are constructed to extract domain-specific information, while the shared encoder extracts the domain-independent information. Subsequently, four loss functions in green squares stand for  $L_{mmd}$ ,  $L_{was-gp}$ ,  $L_{diff-gp}$  and  $L_{cl}$ , which are analyzed in detail in Section 3.2. Finally, the predictions from private classifiers and shared classifier are weighted and summed by the similarity between the source and target features.

Specifically, we sequentially select one subject in the dataset as the target domain data, and the other subjects as the source domain data. As described in [Equation \(1\)](#), let  $\mathbf{X}_S$  be the source data matrices,  $\mathbf{Y}_S$  are their labels, and  $\mathbf{X}_T$  are the unlabeled target data matrices,

$$\mathbf{X}_S = \{\mathbf{X}_S^i\}_{i=1}^n, \mathbf{Y}_S = \{\mathbf{Y}_S^i\}_{i=1}^n, \quad (1)$$

where  $n$  represents the number of subjects in source domains.

In the proposed MISNet, the common encoder  $E_C$  maps the source data matrices  $\mathbf{X}_S^i$  and target data matrices  $\mathbf{X}_T$  to the low-level feature space as shown in [Equation \(2\)](#),

$$\mathbf{X}_S^{\prime i} = E_C(\mathbf{X}_S^i), \mathbf{X}_T^{\prime} = E_C(\mathbf{X}_T). \quad (2)$$

Then for the low-level features of each source domain, we construct a private encoder to obtain its source domain-specific characteristics. For the target domain, the  $n$  private encoders extract domain-specific features as shown in [Equation \(3\)](#),

$$\mathbf{F}_{SP}^i = E_P^i(\mathbf{X}_S^{\prime i}), \mathbf{F}_{TP}^i = E_P^i(\mathbf{X}_T^{\prime}), i = 1, 2, \dots, n, \quad (3)$$

where  $n$  is the number of source domains. Meanwhile as shown in [Equation \(4\)](#), the shared encoder  $E_S$  maps the low-level features of source and target domains to the shared domain,

$$\mathbf{F}_{SS}^i = E_S(\mathbf{X}_S^{\prime i}), \mathbf{F}_{TS} = E_S(\mathbf{X}_T^{\prime}). \quad (4)$$

Subsequently as described in [Equation \(5\)](#), the private classifier  $C_P^i$  and shared classifier  $C_S$  take  $(\mathbf{F}_{SP}^i, \mathbf{F}_{TP}^i)$  and  $(\mathbf{F}_{SS}^i, \mathbf{F}_{TS})$  as the inputs and output the emotion predictions  $(\hat{\mathbf{Y}}_{SP}^i, \hat{\mathbf{Y}}_{TP}^i)$  and  $(\hat{\mathbf{Y}}_{SS}^i, \hat{\mathbf{Y}}_{TS})$ , respectively,

$$\begin{aligned} \hat{\mathbf{Y}}_{SP}^i &= C_P^i(\mathbf{F}_{SP}^i), \hat{\mathbf{Y}}_{TP}^i = C_P^i(\mathbf{F}_{TP}^i), \\ \hat{\mathbf{Y}}_{SS}^i &= C_S(\mathbf{F}_{SS}^i), \hat{\mathbf{Y}}_{TS} = C_S(\mathbf{F}_{TS}). \end{aligned} \quad (5)$$

Finally,  $\hat{\mathbf{Y}}_S$  is the weighted sum of  $\hat{\mathbf{Y}}_{SP}^i$  and  $\hat{\mathbf{Y}}_{SS}^i$ ,  $\hat{\mathbf{Y}}_T$  is that of  $\{\hat{\mathbf{Y}}_{TP}^i\}_{i=1}^n$  and  $\hat{\mathbf{Y}}_{TS}$ , respectively. We use  $\hat{\mathbf{Y}}_S$  to calculate the classification loss  $L_{cl}$  in the training phase and  $\hat{\mathbf{Y}}_T$  to predict the emotion category in the test phase.

#### 3.2 Modules

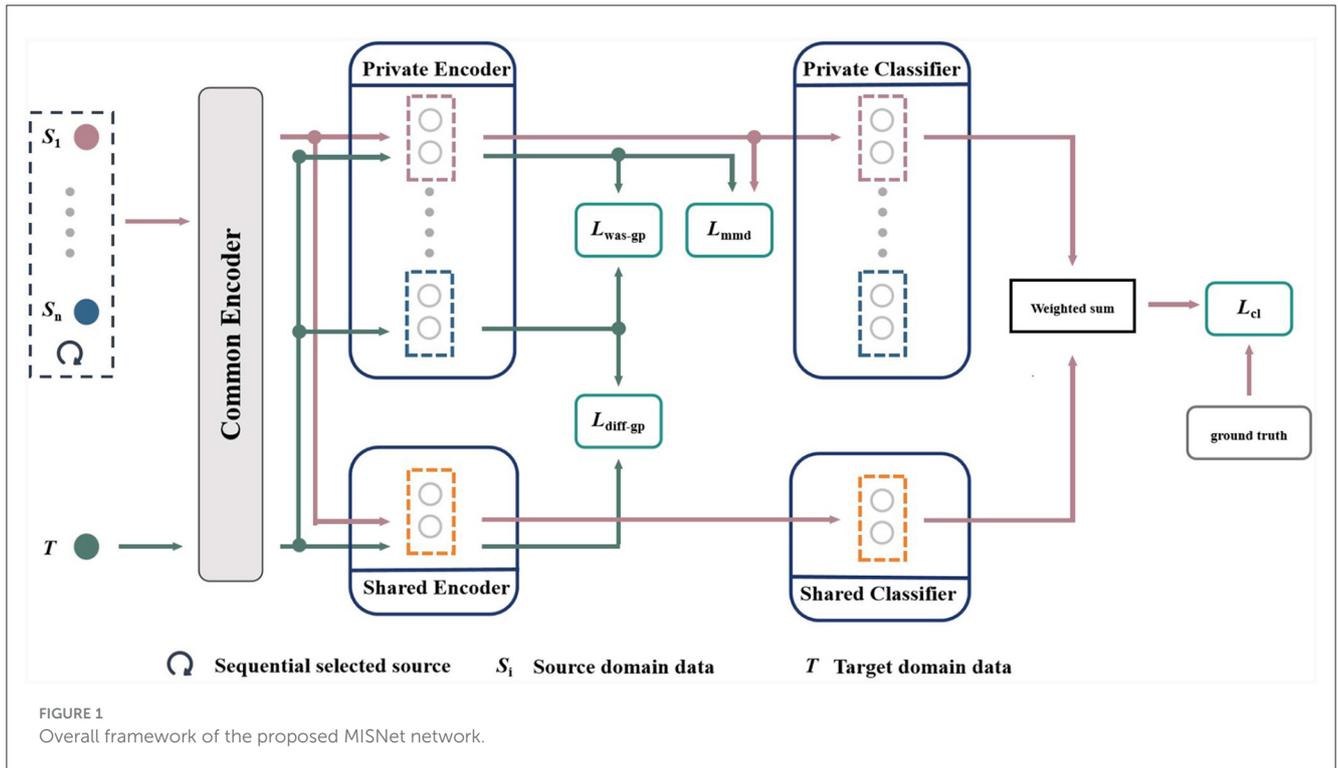
For the domain shift problem in subject independent EEG emotion recognition, we propose to design an EEG emotion recognition model based on feature disentanglement, extracting domain-specific and domain-shared features to improve the robustness and interpretability of the network. A common encoder is used to extract the low-level features of the EEG signal, and the private encoders map the sample data of each domain to its domain-specific features, reducing the distance between the source and target domains after their feature mapping. The shared encoder extracts domain-shared features and imposes secondary constraints on the mapping distance between the source domain and target domain features. The private classifiers and shared classifier map the domain-specific and domain-shared features to predict the emotions of EEG signals. This section will specifically explain the role of each module in the proposed model and the overall optimization strategy.

##### 3.2.1 Common encoder

Despite the individual differences in EEG signals, some common EEG characteristics still exist in the signals of human brain activity. We assume that EEG signals from different subjects share the same shallow feature. Similar to MS-MDA ([Chen et al., 2021a](#)) and MEER-Net ([Chen et al., 2021b](#)), a common encoder maps all domain data into a common latent space, extracting the low-level features of source and target domains. The common encoder is designed to perform the nonlinear mapping of DE features of EEG signals, which obtains the preliminary mapping of emotional information by extracting low-level features from EEG signals. This lays a solid foundation for extracting the individual difference and group commonality of multi-source domains, therefore enhancing the classification performance.

##### 3.2.2 Private encoders and shared encoder

To capture the domain-specific information and consider the difference among different domains, we set up  $n$  fully-connected layers as private encoder for each source domain to map the data from the common feature space to the latent private feature space. Inspired by the idea of feature disentanglement in domain generalization ([Wang and Chen, 2021](#)), the shared emotional information is extracted through the shared encoder by mapping



the low-level feature space to the shared feature space. Note that the shared encoder has the same structure as the private encoder in order to balance their learning abilities. For one iteration in each epoch, the private and shared encoders capture only the features of source domain and target domain that are currently trained. We employ the maximum mean discrepancy (MMD) to calculate the marginal distribution between the source and target domains in reproducing the kernel hilbert space  $\mathcal{H}$ . MMD is often used to measure the distance between two distributions and is a commonly used loss function in transfer learning. The definition of MMD is

$$MMD(f, p, q) = \sup_{\|f\|_{\mathcal{H}} \leq 1} E_p[f(x)] - E_q[f(y)], \quad (6)$$

where  $f$  is the mapping function which is the norm in the reproducing kernel hilbert space. The distributions of  $x$  and  $y$  is  $p$  and  $q$ , respectively, and  $E$  is the mathematical expectation. However, this equation is challenging to calculate because the feature space of  $f$  has infinite dimensions. Thus, Equation 6 is solved by using the linear kernel function to simplify calculation,

$$L_{mmd} = |\psi(\mathbf{F}_{SP}^i) - \psi(\mathbf{F}_{TP}^i)|_{\mathcal{H}} = \overline{(\mathbf{F}_{SP}^i - \mathbf{F}_{TP}^i) * (\mathbf{F}_{SP}^i - \mathbf{F}_{TP}^i)^T}, \quad (7)$$

where  $\psi$  denotes the mapping function, the symbol  $*$  is the matrix multiplication, and  $\bar{x}$  represents the mean of  $x$  in feature dimension.  $L_{mmd}$  dominates the domain adaptation direction, alleviating the feature distribution difference between the source and target domains.

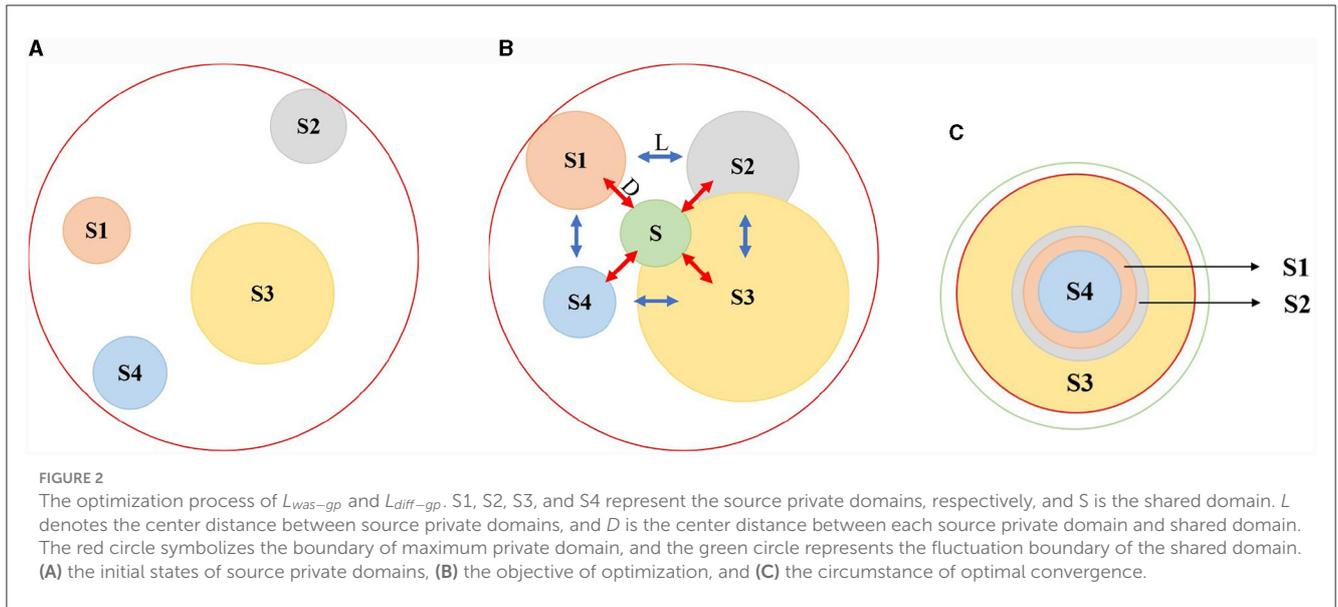
Due to individual differences, all source domains are linearly independent, indicating that their private feature distributions may be quite distinct. This results in a larger spacing among all

source private domains, forming a larger outer contour boundary denoted by the red circle, as shown in Figure 2A. In the process of optimizing iteration, in addition to reducing the distribution distance between the source private and target private domains, it is also necessary to shrink the spacing among source private domains, thereby obtaining a more compact set of source private domains. And thus, the overall boundary of source private domains is also reduced, denoted by the red circle as shown in Figure 2B. On the other hand in Figure 2B, the distribution distance between the shared and private domains is also reduced, forcing the network to extract the domain-independent features. In order to improve the training speed and reduce the network complexity, the above operations are not performed on the distribution distance of different domains, but on the center of each domain.

Specifically, to align the private domains and shared domain, we design two auxiliary loss functions  $L_{was-gp}$  and  $L_{diff-gp}$ . In the current  $i$ -th iteration, the first order  $L_{was}$  is proposed to align the marginal distributions of each private domain as shown in Equation (8),

$$L_{was} = \sum_{j=1, j \neq i}^n |\overline{\mathbf{F}_{TP}^i} - \overline{\mathbf{F}_{TP}^j}|, \quad (8)$$

where  $\overline{\mathbf{F}_{TP}^i}$  and  $\overline{\mathbf{F}_{TP}^j}$  denote the mean vector across feature dimensions of the domain-specific features extracted by the  $i$ -th and  $j$ -th private encoders, respectively. Considering the individual differences and potential outliers of the source domains, we select the features of the target private domain to compress the different private domains through forcing the private encoders to extract the domain-specific information from the target domain rather than the source domains. Furthermore, the soft version of the constraint



with a penalty  $L_{was-gp}$  is enforced on the gradient norm of random samples  $\mathbf{x}_T^i \in \mathbf{X}_T$  to improve the stability of  $L_{was}$  and reduce optimization errors caused by the outlier gradients,

$$L_{was-gp} = \underbrace{\sum_{j=1, j \neq i}^n \frac{|\overline{\mathbf{F}_{TP}^i} - \overline{\mathbf{F}_{TP}^j}|}{L_{was}}}_{\text{gradient penalty}} + \underbrace{(\|\nabla_{\hat{\mathbf{x}}_w} E_p^i(\hat{\mathbf{X}}_w)\|_2 - 1)^2}_{\text{gradient penalty}}, \quad (9)$$

where  $\hat{\mathbf{X}}_w$  is uniformly defined along straight lines between pairs of points sampled from the  $i$ -th target domain-specific feature  $\mathbf{F}_{TP}^i$  and  $j$ -th target domain-specific feature  $\mathbf{F}_{TP}^j$ . This idea is motivated by the WGAN-GP (Gulrajani et al., 2017), where the gradient penalty also adopts the no-batch normalization and two-sided penalty strategy. Different from WGAN-GP,  $\mathbf{F}_{TP}^i$  and  $\mathbf{F}_{TP}^j$  are extracted from different private encoders that have the same input feature  $\mathbf{X}'_T$ . In addition, we also propose  $L_{diff-gp}$  to align the distributions of  $i$ -th private encoder and the shared encoder,

$$L_{diff-gp} = \underbrace{|\overline{\mathbf{F}_{TP}^i} - \overline{\mathbf{F}_{TS}}|}_{L_{diff}} + \underbrace{(\|\nabla_{\hat{\mathbf{x}}_d} E_p^i(\hat{\mathbf{X}}_d)\|_2 - 1)^2}_{\text{gradient penalty}}, \quad (10)$$

where  $\hat{\mathbf{X}}_d$  is calculated similarly as  $\hat{\mathbf{X}}_w$  in Equation (9) by using the target private feature  $\mathbf{F}_{TP}^i$  and the target shared feature  $\mathbf{F}_{TS}$  in the  $i$ -th iteration.

With the progress of loop iteration, the overall boundary of source private domains will shrink rapidly under the constraints of  $L_{was-gp}$  and  $L_{diff-gp}$ . Given an optimal convergence as illustrated in Figure 2C, the centers of the private and shared domains approach one another. The optimal overall private domain has the smallest boundary, which is equal to the boundary of maximum private domain denoted by the red circle as shown in Figure 2C. Moreover, the spacing among the optimal private domains is also minimized or even disappeared. Meanwhile, the fluctuation boundary of the shared domain will approach the boundary of maximum private

domain represented by the green and red circles as illustrated in Figure 2C. Since  $L_{cl}$  and  $L_{mmd}$  dominate the classification and domain adaptation tasks, respectively, the fluctuation boundary of the shared domain cannot easily converge to the optimum result. To sum up, the final convergence of the shared domain center meets the following three requirements:

- Meet the minimum  $L_{cl}$  requirements for the shared domain.
- After the private encoder mapping process, the source and target domains must have the minimum  $L_{mmd}$ .
- Meet the minimum distance requirement between the fluctuation boundary of shared domain and the optimum boundary of private domains.

Furthermore, when there is a conflict during the optimization of the shared domain distribution and  $L_{cl}$  or  $L_{mmd}$ ,  $L_{cl}$  and  $L_{mmd}$  will prioritize to optimize, resulting a small spacing between the boundary of maximum private domain and that of shared domain. Under ideal circumstances, the boundary of the shared domain can be optimal, that is, being the boundary of maximum private domain. At this point, the extracted domain-independent features are optimum for emotion prediction, represented by the overlap of green and red circles in Figure 2C.

### 3.2.3 Private and shared classifiers

Following the private encoders, the private classifiers predict emotion states by using the private features. The softmax activate function is implemented after the fully-connected layer corresponding to each source domain, which transforms hidden states to predict the category label. Like the private classifiers, the shared classifier has the same structure to balance their classification abilities. During the training process, we measure  $L_{cl}$  of private and shared classifiers using the label smoothing cross-entropy loss as described in Equations (11) and (12),

$$L_{cl_p} = - \sum_c^K q(y, c) \log P(c|\hat{Y}_{SP}^i), \tag{11}$$

$$L_{cl_s} = - \sum_c^K q(y, c) \log P(c|\hat{Y}_{SS}),$$

$$q(y, c) = \begin{cases} 1 - \varepsilon & \text{if } c = Y_S, \\ \varepsilon/K - 1 & \text{otherwise,} \end{cases} \tag{12}$$

where  $Y_S$  is the emotion label of the source domain,  $\varepsilon$  is the smooth probability, and  $K$  is the category number of emotions.

### 3.2.4 Weight sum

Considering both the individual differences and group commonalities, we also propose to weight  $L_{cl_p}$  and  $L_{cl_s}$  based on the similarity between the private target domain and private source domain to dynamically adjust the optimization process and balance the weight of the private and shared networks. During the training process, we integrate the private and shared classifiers by calculating MMD between the private and shared domains. The weight of private features  $w_p$  and shared features  $w_s$  is calculated by

$$w_p = |\psi(\mathbf{F}_{SP}^i) - \psi(\mathbf{F}_{TP}^i)|_{\mathcal{H}}, \tag{13}$$

$$w_s = |\psi(\mathbf{F}_{SS}) - \psi(\mathbf{F}_{TS})|_{\mathcal{H}}. \tag{14}$$

And  $L_{cl}$  is calculated by,

$$L_{cl} = \frac{w_s}{w_p + w_s} * L_{cl_p} + \frac{w_p}{w_p + w_s} * L_{cl_s}. \tag{15}$$

The weighted sum minimizes the distance between the private and shared domains. The deeper reason is the dynamic adjustment of the optimization process, as the smaller the MMD between the source and target domains, the smaller the difference in their distributions. Specifically, if the distribution between source and private shared domains are closer,  $w_s$  will be smaller, and then the weight of private encoder is also smaller. Due to the back propagation theory, the less gradient is assigned to private encoder, the more would be assigned to shared one relatively, indicating that the shared encoder has more learning capabilities than private encoder. Therefore, more learning capabilities are assigned to the corresponding encoder. By using Equation (15), the outputs of network are weighted based on their distributions, and thus more attentions are paid to the inter-domain predictions with more similarities.

Given  $L_{cl}$ ,  $L_{mmd}$ ,  $L_{was-gp}$  and  $L_{diff-gp}$ , the final loss function is represented as,

$$L = L_{cl} + \alpha L_{mmd} + \beta L_{was-gp} + \gamma L_{diff-gp}, \tag{16}$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are the hyper-parameters. To sum up,  $L_{cl}$  is the classification loss function, which controls the overall optimization direction of the model.  $L_{mmd}$  is the domain adaptation loss

**Input:** Source domain dataset

$\mathbf{X}_S = \{\mathbf{X}_S^i\}_{i=1}^n$  and ground truth

$\mathbf{Y}_S = \{\mathbf{Y}_S^i\}_{i=1}^n$ , target domain dataset  $\mathbf{X}_T$

**Output:** Predicted emotional state of target domain  $\hat{\mathbf{Y}}_T$

1 Pretraining  $E_C$  and  $E_S$  with subject-dependent emotion classification task

2 Randomly initialize  $E_p^{1\sim n}$ ,  $C_p^{1\sim n}$ , and  $C_s$

3 **Training phase:**

4 **for epoch in epochs do**

5 **for i in n do**

6  $E_C(\mathbf{X}_S) \rightarrow \mathbf{X}'_S$ ,  $E_C(\mathbf{X}_T) \rightarrow \mathbf{X}'_T$

7  $E_p^i(\mathbf{X}'_S) \rightarrow \mathbf{F}_{SP}^i$ ,  $E_p^i(\mathbf{X}'_T) \rightarrow \mathbf{F}_{TP}^i$

8  $(\mathbf{F}_{SP}^i, \mathbf{F}_{TP}^i) \rightarrow$  Equation (7),

9  $(\mathbf{F}_{TP}^i, \mathbf{F}_{TP}^i)^{1\sim n} \rightarrow$  Equation (9)

10  $\theta_{E_C} \leftarrow \theta_{E_C} + lr * \frac{\partial(L_{mmd} + L_{was-gp})}{\partial \theta_{E_C}}$ ,

11  $\theta_{E_p^i} \leftarrow \theta_{E_p^i} + lr * \frac{\partial(L_{mmd} + L_{was-gp})}{\partial \theta_{E_p^i}}$

12  $E_S(\mathbf{X}'_S) \rightarrow \mathbf{F}_{SS}$ ,  $E_S(\mathbf{X}'_T) \rightarrow \mathbf{F}_{TS}$

13  $(\mathbf{F}_{TP}^i, \mathbf{F}_{TS}) \rightarrow$  Equation (10)

14  $(\mathbf{F}_{SP}^i, \mathbf{F}_{TP}^i) \rightarrow$  Equation (13)

15  $(\mathbf{F}_{SS}, \mathbf{F}_{TS}) \rightarrow$  Equation (14)

16  $C_p^i(\mathbf{F}_{SP}^i)$ ,  $C_s(\mathbf{F}_{SS}) \rightarrow$  Equation (15)

17  $\theta_{E_C} \leftarrow \theta_{E_C} + lr * \frac{\partial L_{cl}}{\partial \theta_{E_C}}$ ,

18  $\theta_{E_p^i} \leftarrow \theta_{E_p^i} + lr * \frac{\partial(L_{cl_p} + L_{diff-gp})}{\partial \theta_{E_p^i}}$ ,

19  $\theta_{E_S} \leftarrow \theta_{E_S} + lr * \frac{\partial(L_{cl_s} + L_{diff-gp})}{\partial \theta_{E_S}}$

20 Update network by minimizing the total loss  $\rightarrow$  Equation (16)

21 **end**

22 **Test phase:**

23  $E_p^{1\sim n}(E_C(\mathbf{X}_T)) \rightarrow \mathbf{F}_{TP}^{1\sim n}$ ,  $E_S(E_C(\mathbf{X}_T)) \rightarrow \mathbf{F}_{TS}$

24 Predict target labels  $(\mathbf{F}_{TP}^{1\sim n}, \mathbf{F}_{TS}) \rightarrow$  Equation (17)

25 **Return:**  $\hat{\mathbf{Y}}_T$

Algorithm 1. Workflow of the proposed MISNet framework.

function, which controls the domain adaptation direction of the model.  $L_{was-gp}$  aligns the marginal distributions of each private domain.  $L_{diff-gp}$  aligns the distributions of  $i$ -th private encoder and the shared encoder. By combing those four loss functions as in Equation (16), the model can obtain the ability of alleviating negative transfer by considering the individual difference and group commonality simultaneously.

In the test phase, we assume that the optimal convergence boundary has been reached, and the predictions of the private and shared domains are added together to output the final results,

$$\hat{\mathbf{Y}}_T = \frac{1}{n} \sum_{i=1}^n (C_p^i(\mathbf{F}_{TP}^i) + C_s(\mathbf{F}_{TS})). \tag{17}$$

In summary, the workflow of the proposed MISNet is presented in Algorithm 1.

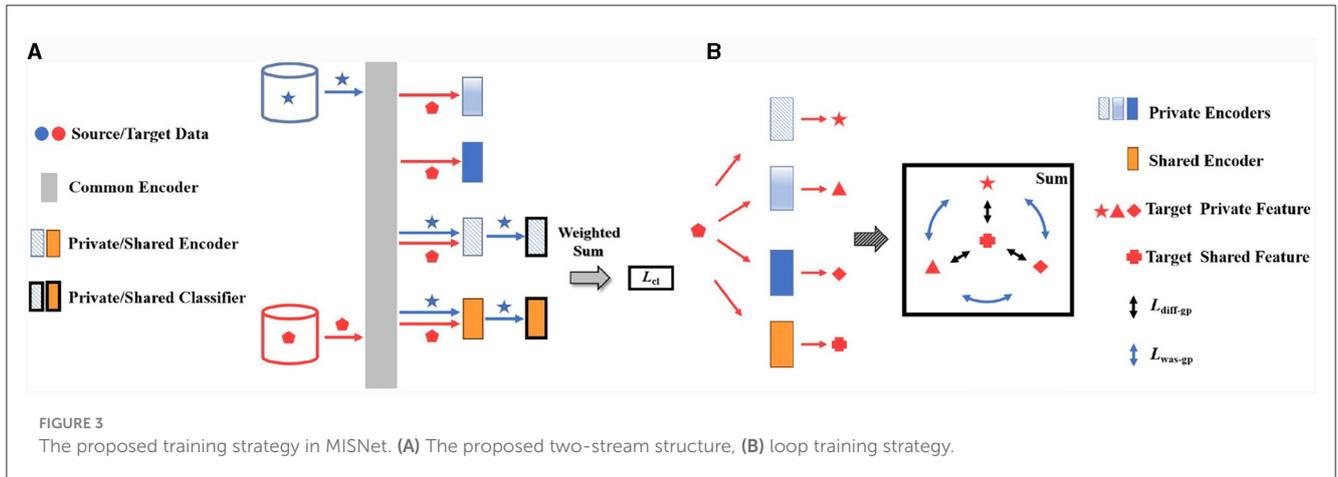


FIGURE 3 The proposed training strategy in MISNet. (A) The proposed two-stream structure, (B) loop training strategy.

### 3.3 Training strategy

Since the main convergence direction of different private encoders is determined by the classification loss  $L_{cl}$  of source domain and their initial feature mapping rules are determined by the distributions of source domains, when there is an outlier source domain, it may occur that one private domain mapping of target domain is far from other private domain mappings. Consider an extreme case of  $L_{was-gp}$  in Equation (9), when a private feature  $F_{TP}^i$  is far from other private features  $F_{TP}^j$  ( $j \neq i$ ) in the reproducing kernel hilbert space. If directly using the traditional parallel training strategy, this outlier will affect other private encoders and classifiers due to the addition operation of loss function. Therefore, the prediction and mapping rules of private encoder in the outlier source domain is significantly different from other private encoders, resulting in a large difference between  $F_{TP}^i$  and  $F_{TP}^j$ , ultimately impacting the training process.

In the above case,  $L_{diff-gp}$  in Equation (10) will also be somewhat affected. Since one or more source domains may be distant from other sources, the mapping rules will have large differences even when inputting the same low-level feature of the target domain. Additionally, the model training deviation caused by the outlier source domains is mixed in the shared domain, misguiding the optimization direction of shared encoder and the convergence boundary of shared classifier. The distance between the private domain mapping of outlier source and the shared domain will inevitably cause the outlier domain mapping to stay away from other source domain mappings. In this case of improper optimization, the model tends to converge to most source domains while ignoring the outlier domain, the boundary of shared domain shifts toward the concentrated source domains and deviates from that of maximum private domain, resulting the model not to converge.

In order to alleviate the outlier source problem, we propose the two-stream training structure instead of the parallel one, by only inputting the data of current source domain and unlabeled data of target domain during each iteration. The proposed two-stream structure is depicted in Figure 3A. By separating different source domains in the training process,  $L_{was-gp}$  is calculated between the current target private feature  $F_{TP}^i$  and others  $F_{TP}^j$  ( $j \neq i$ ), alleviating

the effects caused by the other source domains. Similarly,  $L_{diff-gp}$  only calculates the distance between the target feature of current private encoder and feature of the shared encoder, avoiding mixing the outlier source domain. As a result, the model errors caused by outlier source domain would be further alleviated.

With the proposed two-stream training structure, we adopt the loop iteration strategy as illustrated in Figure 3B to perform the sequential iterations of all source domains, alleviating domain confusion in loss function and improving the robustness of the network. Here, we refer to a loop of the whole source domains as an epoch comprising  $n$  iterations, where an iteration corresponds one source domain. Specifically during each iteration, the proposed MISNet successively selects the current source data matrix  $X_S^i$  and its corresponding encoder  $E_p^i$  as the private encoder, then  $L_{mmd}$  measures the marginal distribution between  $F_{SP}^i$  and  $F_{TP}^i$ , and  $L_{was-gp}$  aligns the current target private feature  $F_{TP}^i$  and others  $F_{TP}^j$  presented as the blue double arrows in Figure 3B. Meanwhile,  $L_{diff-gp}$  aligns the distribution between the current target private feature  $F_{TP}^i$  and target shared feature  $F_{TS}$  presented as the black double arrows in Figure 3B. During the sequential iterations, the change of source domains in two adjacent iterations will cause slight fluctuations in  $L_{was-gp}$  and  $L_{diff-gp}$ . As the loop iteration progresses, the total loss of the same iteration will decrease between two adjacent loops until it converges.

To make maximum use of the source domain data and ensure that the initial mapping of shared encoder contains sufficient emotional information, we pre-train the common encoder and shared encoder in a subject-dependent emotion classification task. All shuffled source domain data and their labels are used for the training. The common encoder firstly extracts the low-level features, and then the shared encoder is used to capture the deep features. Finally, the shared classifier calculates cross-entropy loss between the output and ground truth labels. In addition, the Adam optimizer is used as the optimizing function, the total epoch is set to 100 and batch size is set to 64. After the pre-training, only the weights of common encoder and shared encoder are saved during the training phase by employing a normally initialized classifier, so that the encoders are initialized to reasonable parameters and the model convergence would be accelerated, which avoids random

initialization causing the model to not converge or converge to local optima.

We can also re-examine the network construction from the perspective of loss function design by using the following two criteria:

- $L_{mmd}$  only reduces the distance between the source and target domains in the private domain instead of using the shared domain.
- The auxiliary losses of  $L_{was-gp}$  and  $L_{diff-gp}$  use the features of target domain to narrow the corresponding encoder mapping instead of those of source domain.

The purpose of the first criterion is to avoid misleading the direction of domain adaptation when the shared encoder  $E_S$  is trained mainly by the classification loss of  $L_{cl}$  which extracts the target domain information containing the individual differences among the source domains.

For the  $L_{was-gp}$  loss in the second criterion, if directly using a single source domain  $\mathbf{X}_{SP}^i$  as the input of private encoders and aligning their outputs in each iteration, the wrong gradient of domain information in  $\mathbf{X}_S^i$  would be mixed into other encoders  $\mathbf{F}_{TP}^j$  ( $j \neq i$ ). Therefore, we use the target private features  $\mathbf{F}_{TP}^i$  and  $\mathbf{F}_{TP}^j$  to align different private domains in  $L_{was-gp}$ . On the other hand, in the actual iteration process of  $L_{diff-gp}$ , each private encoder is simultaneously updated by  $L_{cl}$  and  $L_{mmd}$ , and different source domain is input in sequence during one loop. When directly adopting the features of source domain as the input will introduce fluctuations in  $L_{diff-gp}$ , leading to the training collapse. Therefore, we use the features of target domain instead to constrain the domain adaptation direction. After aligning the target private features  $\mathbf{F}_{TP}^i$  and target shared features  $\mathbf{F}_{TS}$ , the shared classifier  $C_S$  could classify the emotion of target domain correctly.

## 4 Experimental settings

This section describes the datasets used for evaluation, EEG data pre-processing and implementation details in the proposed MISNet.

### 4.1 Datasets

We evaluate the proposed network on SEED (Duan et al., 2013; Zheng and Lu, 2015; Liu et al., 2022) and SEED-IV (Zheng et al., 2019), which are public datasets commonly used for EEG emotion recognition. The SEED dataset contains EEG signals from 15 Chinese participants (seven males and eight females). The participants are required to watch 15 Chinese film clips chosen from a pool of materials as stimuli to elicit positive, neutral and negative emotion. Additionally, each film clip contains scenes and audios that is ~4-min long to prevent viewer fatigue. The clip order prevents the continuous display of two clips depicting the same emotion category. Each subject participates in three sessions containing 15 trials, where each session is conducted on a separate day. For feedback purposes, the participants are asked to complete a questionnaire and report their emotional responses immediately

after viewing each clip. The EEG signals are recorded by an ESI NeuroScan system at a sampling rate of 1,000 Hz through a 62-electrode cap according to the international 10-20 system. The SEED-IV dataset is similar to the SEED, but it has four emotion categories (happiness, sadness, fear and neutral) and conducts 24 trials per session.

### 4.2 Data pre-processing

For the data pre-processing of EEG signals, the original EEG data was downsampled to 200 Hz and a bandpass filter from 0 to 75 Hz was applied, and a 512-point short-time Fourier transform was used with a non-overlapped Hanning window of 1 s to calculate the frequency domain features. Considering their effectiveness in the EEG emotion recognition task (Yang et al., 2017; Li et al., 2022), the DE features were then computed on the five bands as  $\delta$ : 1–3 Hz,  $\theta$ : 4–7 Hz,  $\alpha$ : 8–13 Hz,  $\beta$ : 14–30 Hz and  $\gamma$ : 31–50 Hz (Zheng and Lu, 2015). For the gaussian distribution, the DE feature is defined as shown in Equation (18),

$$\begin{aligned} \text{DE}(\mathbf{X}) &= - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{2} \log 2\pi e\sigma^2, \end{aligned} \quad (18)$$

where  $\mathbf{X}$  obeys the gaussian distribution  $N(\mu, \sigma^2)$ ,  $x$  is the element of  $\mathbf{X}$ . Therefore, the 310-dimensional DE features (62 channels multiplying with five frequency bands) were computed, and the features were smoothed with the conventional moving average and linear dynamic system. After the pre-processing steps, each session contains 3,394 samples for the SEED dataset and 822 samples for SEED-IV dataset.

### 4.3 Implementation details

In the proposed MISNet, the common encoder is a 3-layer fully-connected layer with 310-256-128-64 nodes, which extracts the low-level features of source and target domains. Each private encoder and shared encoder are composed of a fully connected layer designed as 64 (input layer)—32 (output layer)—LeakyRelu activation. Besides, a single fully-connected layer is chosen for each private classifier and shared classifier with a hidden dimension from 32 to the number of emotion categories. Note that there is no batch normalization layer, since we use the gradient penalty guidelines in Equations (9) and (10). The LeakyRelu activation function with a negative slope of 0.01 is used in all hidden layers. In addition, we normalize the data of source and target domains to enhance performance by using the electrode-wise method in Chen et al. (2021a).

For the hyper-parameters in Equation (16), we consider the trade-off among the primary losses of  $L_{cl}$  and  $L_{mmd}$  as well as auxiliary effects of  $L_{was-gp}$  and  $L_{diff-gp}$ , and set  $\alpha = \frac{2}{1+e^{-10 \times k / (nos \times epoch)}} - 1$  ( $k = 1, 2, \dots, nos \times epoch$ ), where  $nos$  means number of samples and  $\beta = \gamma = \frac{\alpha}{100}$ . In our network, we set the learning rate to 0.01 and the batch size to 64. In addition, the

Adam optimizer is used as the optimizing function, the total epoch is set to 200, and a cosine annealing schedule is used to determine the learning rate for each epoch. The proposed framework is implemented in PyTorch with version of 1.11 on NVIDIA RTX 1080Ti GPU. The model parameters and computation cost are FLOPs 9.88M and params 154.4K, respectively.

## 5 Experiments and results

In this section, we first test different values of loss weights  $\beta$  and  $\gamma$  to evaluate the effectiveness of the hyper-parameter settings. Next, the ablation experiments are conducted in terms of loss functions, dynamic convergence of network and visualization of mapping features. Then, we compare the proposed network with other competing methods by using the LOSO strategy on the SEED and SEED-IV datasets. Finally, the experiments of adding noise are conducted to demonstrate the robustness of the proposed network.

### 5.1 Hyper-parameter evaluation

In this section, we evaluate the hyper-parameters of  $\beta$  and  $\gamma$  in Equation (16) within a certain range based on previous experience to explore the impact of different settings in terms of accuracy on the SEED dataset. The corresponding results are illustrated in Figure 4. It can be seen from Figure 4 that, both  $L_{was-gp}$  and  $L_{diff-gp}$  are affected by the selected hyper-parameters of  $\beta$  and  $\gamma$ , with  $L_{diff-gp}$  being more sensitive than  $L_{was-gp}$ . Even with the worst recognition result of 83.05% by setting  $\beta = \frac{\alpha}{100}$  and  $\gamma = \frac{\alpha}{1000}$ , the proposed network still maintains the strong ability of subject independent emotion recognition. Since setting  $\beta = \gamma = \frac{\alpha}{100}$  has achieved best recognition performance on SEED dataset, we use this setting to evaluate the effectiveness of network in the following experiments.

### 5.2 Ablation study

To demonstrate the effectiveness of loss functions in MISNet, we evaluate the performance of the ablated network on the SEED and SEED-IV datasets, as shown in Table 1.

The subject independent recognition performance is evaluated by using the metrics of mean accuracy (Mean) and standard deviation (Std.). Table 1 indicates that all loss functions can improve recognition performance, affording the mean accuracies of 88.80 and 74.60% on the SEED and SEED-IV datasets, respectively. In addition, the proposed MISNet achieves the standard deviation of 6.24 and 9.30% on the SEED and SEED-IV datasets, respectively, showing better inter-subject stability. Discarding  $L_{mmd}$  in our framework leads to a significant performance degradation compared with depriving other loss functions, proving its importance in domain adaptation. And thus, the higher weight should be assigned to the loss of  $L_{mmd}$  than those of  $L_{was-gp}$  and  $L_{diff-gp}$ . Furthermore, removing  $L_{was-gp}$  and  $L_{diff-gp}$  simultaneously will damage the domain adaptability more than removing any of them individually, since they control the relationship between private features and shared features jointly.

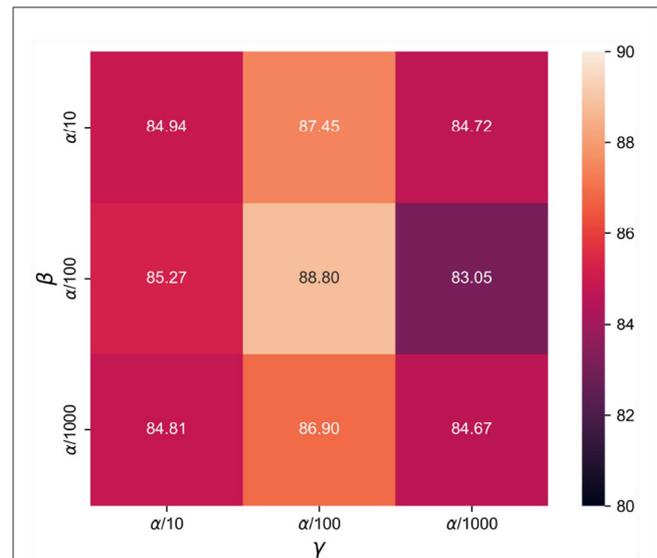


FIGURE 4 Evaluations of different settings of  $\beta$  and  $\gamma$  in terms of accuracy on the SEED dataset.

TABLE 1 Ablation study of loss functions on SEED and SEED-IV datasets.

Variants	SEED		SEED-IV	
	Mean	Std.	Mean	Std.
MISNet	88.80	6.24	74.60	9.30
w/o* gradient penalty	86.52	8.42	67.12	12.65
w/o $L_{mmd}$	60.47	8.48	56.62	14.22
w/o $L_{was-gp}$	86.72	8.91	70.56	13.15
w/o $L_{diff-gp}$	87.36	10.48	69.50	13.72
$L_{cl}$ and $L_{mmd}$	82.71	15.25	67.64	12.26
only $L_{cl}$	60.20	11.02	54.78	12.35

\*w/o  $L_*$  denotes the ablated network trained without the loss function  $L_*$ .

In addition, the gradient penalty of two auxiliary loss functions  $L_{was-gp}$  and  $L_{diff-gp}$  allows stable convergence of the private and shared domains, which is reflected in the improvement of average accuracy and decrease of standard deviation.

One of the primary goals of the proposed network is to align the distributions of private and shared domains to alleviate the negative transfer in domain adaptation caused by individual differences in EEG signals. Next, we evaluate the convergence process of proposed MISNet and visualize the domain mapping of the private and shared domains in a two-dimensional way by using the t-distributed stochastic neighbor embedding (t-sne). The t-sne method is employed to evaluate the similarity between feature representations during the training process, meaning the closer points have a higher similarity in real space. The similarity between source and target domains affects the domain adaptability of network, while the distinction among the emotion categories reflects its emotion discrimination. The dynamic convergence process of the proposed MISNet is shown in Figure 5.

It can be depicted from Figure 5A that the original distributions of the DE feature projection of different source domains are chaotic and irregular, due to the individual differences of EEG signals.

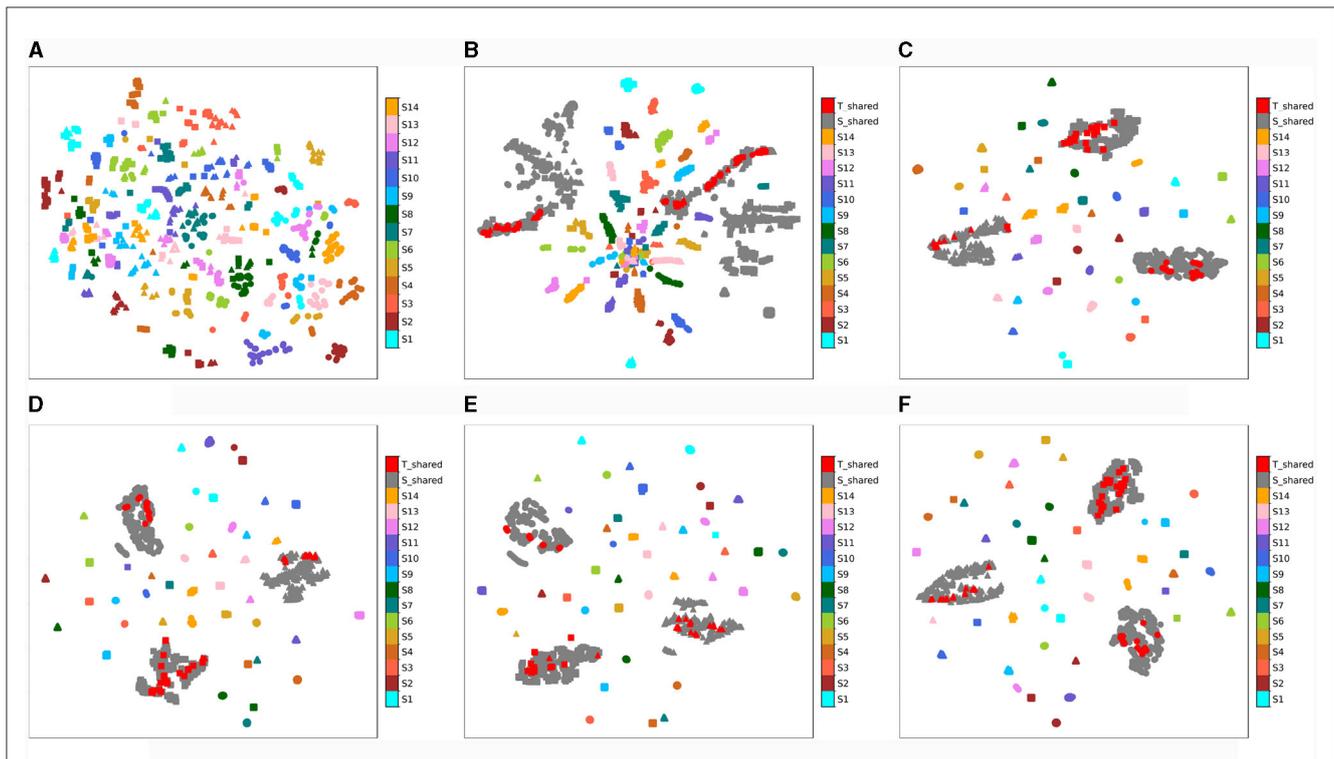


FIGURE 5

Training process of the proposed MISNet with t-sne mapping. Red color denotes the target shared features and gray color represents the source shared features, while other colors symbolize the source private features. Different shapes denote the different emotion categories: ●, ▲, ■ represents positive, neutral and negative emotions, respectively. (A) the original distributions of the DE feature projection of different source domains, (B) the initialization effect of subject-dependent pre-training, (C–F) the training process of MISNet.

After the designed subject-dependent pre-training strategy, the different emotional categories of source shared domain represented by different gray shapes have been distinguished to some extent, which means the shared encoder has a fundamental emotion classification ability, as illustrated in Figure 5B. Figures 5C–F reveal that as the convergence process progresses, the private domain of each source domain (represented by different colors) is clustered based on their emotion category (represented by different shapes), indicating a gradual improvement of emotional discrimination. In addition, the cluster center of each domain is distributed near the middle of the figure, indicating that all cluster centers are aligned among the private domains. As the convergence process progresses, the space of private domains gradually shrinks, indicating that the model is eliminating the interference of spacing among private domains. Furthermore, it can also be found from Figure 5F that, when the model has converged, the source private domains exhibit different emotional distribution patterns, this is because we align the domain centers instead of aligning the distribution of private features in  $L_{was-gp}$ . Additionally, the target shared features (represented by red) are almost always within the source shared features (represented by gray), indicating that the shared encoder can effectively capture the shared emotion information. After the final optimization, the center of the shared domains roughly coincides with that of the private domains, and the boundary of shared domains is close to that of maximum private domain. It can be concluded that the convergence process of MISNet is identical to the anticipated and

has a strong domain adaptability for subject independent EEG emotion recognition.

### 5.3 Comparisons with competing methods

In this section, we compare the proposed MISNet with several competing methods on the SEED and SEED-IV datasets. Table 2 shows the comparison results in terms of the mean classification accuracy and standard deviation with competing methods. Here the results of MS-MDA (Chen et al., 2021a) were obtained by using the LOSO strategy with the open source codes.<sup>1</sup>

It can be seen from Table 2 that the domain adaptation-based methods significantly improve the recognition performance compared with directly using SVM in the subject independent experiments. Furthermore, most domain adaptation methods using multi-source (Chen et al., 2021a,b; Luo and Lu, 2021; Gong et al., 2022; Zhu et al., 2022) can attain better performance than those without multi-source (Li H. et al., 2018; Luo et al., 2018; Ma et al., 2019), which indicates the importance of considering individual differences inside the source domains. Specifically, the proposed MISNet outperforms most of the competing methods, achieving a mean accuracy of 88.8 and 74.6% on the SEED and SEED-IV datasets, respectively. This is attributed to the designed loss functions  $L_{was-gp}$  and  $L_{diff-gp}$ , which enhance the

<sup>1</sup> <https://github.com/VoiceBeer/MS-MDA>

domain adaptability of the network. The evaluation index of standard deviation means the stability performance of networks across subjects in the dataset. Compared with the typical existing methods on two benchmark datasets of SEED and SEED-IV, the proposed MISNet demonstrates the competitive ability overcome the individual differences. Although a small gap exists compared to wMADA- $\beta$  (Luo and Lu, 2021), the proposed method has simpler parameter tuning process during the training phase compared to wMADA- $\beta$  (Luo and Lu, 2021).

In order to prove the generality of the proposed model, we compare it with the competitive method of MS-MDA (Chen et al., 2021a) by using the indicators of F1 score, sensitivity and specificity in Table 3. In the experiment, all folded performance was used to estimate the indicators of F1 score, sensitivity and specificity under the LOSO strategy. It can be seen from Table 3 that, the proposed MISNet has higher generalization ability on all three aspects than MS-MDA (Chen et al., 2021a) and performs stably in all subjects.

To further demonstrate the robustness of the proposed MISNet network, we evaluate the performance of the network on the SEED

dataset by adding gaussian noise in the test data, as shown in Equation (19),

$$\mathbf{X}_T^N = \mathbf{X}_T + \mathbf{K}N, \quad (19)$$

where  $\mathbf{X}_T$  are the target matrices,  $N$  denotes the gaussian noise which obeys the gaussian distribution  $N(0, 1)$  and  $K$  is the noise coefficient. In the experiment, we verify the robustness of the proposed network to noise by gradually increasing the noise coefficient, as shown in Figure 6. It can be seen from Figure 6 that, with the noise coefficient  $K$  increasing from 0 to 0.3, that is, the signal-to-noise ratio gradually decreases, the recognition accuracy of the model gradually decreases inevitably, and the recognition variance increases. When the signal-to-noise ratio is relatively large, that is, when  $K$  is equal to 0.1 and 0.2, the mean accuracy of the proposed model still remains above 75%, indicating that it has a strong tolerance for noise.

In order to show the recognition ability of the proposed MISNet among different emotion categories, Figure 7 shows the confusion matrix on the SEED and SEED-IV datasets. It can be seen from Figure 7 that on the SEED dataset, the MISNet achieves the recognition accuracies of 78.77, 93.02, and 95.03%, respectively on the emotion categories of negative, neutral and positive, demonstrating strong discriminative capability across emotions. And the results on the SEED-IV dataset infer that our framework has decent accuracies for the emotion categories of neutral, sad and happy. While for the category of fear, the proposed MISNet confuses it with sad because these two emotions are relatively similar on EEG signals.

TABLE 2 Comparison results of the proposed MISNet with competing methods on the SEED and SEED-IV datasets.

Methods	SEED		SEED-IV	
	Mean	Std.	Mean	Std.
SVM (Zheng and Lu, 2016)	56.7	4.6	—	—
WGANDA (Luo et al., 2018)	87.1	7.1	—	—
DAN (Li H. et al., 2018)	83.8	8.6	—	—
DResNET (Ma et al., 2019)	85.3	8.0	—	—
PPDA (Zhao et al., 2021)	86.7	7.1	—	—
MS-MDA (Chen et al., 2021a)	82.9	6.4	62.8	10.9
MEER-NET (Chen et al., 2021b)	87.1	2.0	71.0	12.1
wMADA- $\beta$ (Luo and Lu, 2021)	<b>89.3</b>	4.0	—	—
MWACN (Zhu et al., 2022)	87.6	4.0	<b>74.4</b>	—
MISDA (Gong et al., 2022)	<b>88.1</b>	9.5	<b>73.8</b>	11.9
<b>MISNet (Ours)</b>	<b>88.8</b>	6.2	<b>74.6</b>	9.3

The bold font in the table represents the top three in the results.

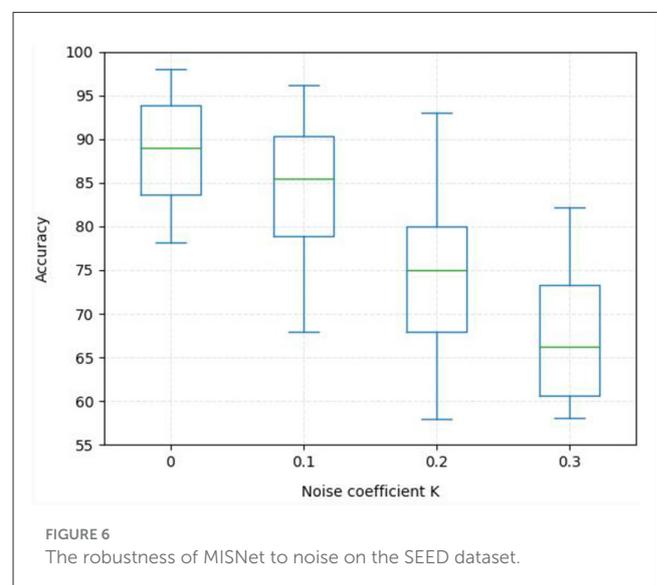


FIGURE 6 The robustness of MISNet to noise on the SEED dataset.

TABLE 3 Comparison generality with competing method on the SEED and SEED-IV datasets.

Variants	SEED			SEED-IV		
	F1-score	Sensitivity	Specificity	F1-score	Sensitivity	Specificity
MS-MDA (Chen et al., 2021a)	79.8 ± 15.1	80.1 ± 15.2	90.3 ± 7.4	42.8 ± 15.4	44.9 ± 14.8	81.8 ± 5.2
MISNet (Ours)	88.7 ± 7.3	88.8 ± 7.2	94.5 ± 3.1	73.2 ± 13.3	74.1 ± 11.9	92.3 ± 3.9

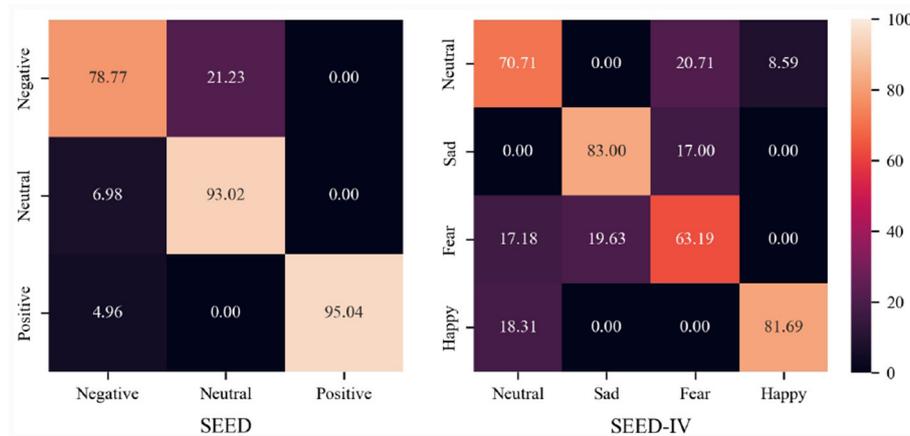


FIGURE 7  
Confusion matrices of predictions on SEED and SEED-IV dataset.

## 6 Conclusion

Although the EEG signals have the advantage of spontaneous and non-subjective characteristic in emotion recognition filed, they still have several limitations, including the individual difference and noisy labeling issues. In this paper, the main challenge we aim to address is the domain shift problem caused by the non-stationary of EEG signals and the individual differences among users.

For the purpose of alleviating the domain shift problem, we propose to consider the individual differences and group commonalities simultaneously, improving the domain adaptation ability of the model. In the proposed MISNet, the decoupling network structure is designed to extract the private domain features and shared domain features of each domain data. In order to constrain overall optimization direction, the classification loss function and domain adaptation loss functions are adopted. In addition, we analyze the convergence process of network to design the auxiliary loss functions of  $L_{was-gp}$  and  $L_{diff-gp}$  in order to align the different domain centers. A pre-training strategy is also used to enhance model stability and ensure that the initial mapping of shared encoder contains sufficient emotional information. Furthermore, the convergence process of the proposed network is dynamically displayed through t-sne mapping. The results on the SEED and SEED-IV datasets demonstrate the effectiveness of our proposed MISNet frameworks.

Since the proposed MISNet needs the unlabeled data of target domain to obtain domain information, it is available for the offline situations in real life, and achieves high-quality emotional awareness by decoupling personality and common emotional characteristics. Our future work will focus on disentangling the domain information from EEG data with a reasonable explanation, thereby constructing a more robust network.

## Data availability statement

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

## Ethics statement

The studies involving humans were approved by Ethics and Morality Committee of Communication University of China. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

MG: Conceptualization, Data curation, Writing—original draft, Formal analysis, Methodology. WZ: Conceptualization, Writing—review & editing, Formal analysis, Funding acquisition. LY: Investigation, Writing—review & editing, Validation, Visualization. QZ: Funding acquisition, Writing—review & editing, Supervision, Project administration.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (No. 62271455)

and the Fundamental Research Funds for the Central Universities (No. CUC18LG024).

## 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.

## References

- Blanchard, G., Lee, G., and Scott, C. (2011). "Generalizing from several related classification tasks to a new unlabeled sample," in *Advances in Neural Information Processing Systems*, Vol. 24 (La Jolla, CA), 1–9.
- Blitzer, J., McDonald, R., and Pereira, F. (2006). "Domain adaptation with structural correspondence learning," in *The Conference on Empirical Methods in Natural Language Processing* (Stroudsburg, PA), 120–128.
- Cai, H., Qu, Z., Li, Z., Zhang, Y., Hu, X., and Hu, B. (2020). Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Inf. Fus.* 59, 127–138. doi: 10.1016/j.inffus.2020.01.008
- Chen, H., Jin, M., Li, Z., Fan, C., Li, J., and He, H. (2021a). 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
- Chen, H., Li, Z., Jin, M., and Li, J. (2021b). "MEERNet: multi-source EEG-based emotion recognition network for generalization across subjects and sessions," in *The 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (Los Alamitos, CA), 6094–6097.
- Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.* 16:031001. doi: 10.1088/1741-2552/ab0ab5
- Cui, Y., Xu, Y., and Wu, D. (2019). EEG-based driver drowsiness estimation using feature weighted episodic training. *IEEE Transact. Neural Systems. Rehabil. Eng.* 27, 2263–2273. doi: 10.1109/TNSRE.2019.2945794
- Ding, Y., Robinson, N., Zeng, Q., Chen, D., Wai, A. A. P., Lee, T.-S., et al. (2020). "Tseption: a deep learning framework for emotion detection using EEG," in *International Joint Conference on Neural Networks* (Oxon: IEEE), 1–7.
- Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194. doi: 10.1126/science.1076358
- Duan, R.-N., Zhu, J.-Y., and Lu, B.-L. (2013). "Differential entropy feature for EEG-based emotion classification," in *The 6th International IEEE/EMBS Conference on Neural Engineering* (Los Alamitos, CA), 81–84.
- Gong, M., Zhong, W., Hu, J., Ye, L., and Zhang, Q. (2022). "Multi-source information-shared domain adaptation for EEG emotion recognition," in *The 5th Chinese Conference on Pattern Recognition and Computer Vision, Shenzhen, China, November 4-7, Part II* (Berlin: Springer), 441–453.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein GANs. *Adv. Neural Inf. Process. Syst.* 30.
- Gupta, V., Chopda, M. D., and Pachori, R. B. (2018). Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals. *IEEE Sens. J.* 19, 2266–2274. doi: 10.1109/JSEN.2018.2883497
- Jiang, Y., Zhang, Y., Lin, C., Wu, D., and Lin, C.-T. (2020). EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system. *IEEE Transact. Intell. Transport. Syst.* 22, 1752–1764. doi: 10.1109/TITS.2020.2973673
- Kim, J., and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, 2067–2083. doi: 10.1109/TPAMI.2008.26
- Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors* 18:401. doi: 10.3390/s18020401
- 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* (Berlin: Springer), 403–413.
- Li, J., Zhao, J., and Lu, K. (2016). "Joint feature selection and structure preservation for domain adaptation," in *The Twenty-Fifth International Joint Conference on Artificial Intelligence* (Palo Alto, CA: AAAI Press), 1697–1703.
- Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., et al. (2022). EEG based emotion recognition: a tutorial and review. *ACM Comput. Surv.* 55, 1–57. doi: 10.1145/3524499
- Li, Y., Zheng, W., Zong, Y., Cui, Z., Zhang, T., and Zhou, X. (2018). A bi-hemisphere domain adversarial neural network model for EEG emotion recognition. *IEEE Transact. Affect. Comp.* 12, 494–504. doi: 10.1109/TAFFC.2018.2885474
- Liu, W., Qiu, J.-L., Zheng, W.-L., and Lu, B.-L. (2022). Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transact. Cogn. Dev. Syst.* 14, 715–729. doi: 10.1109/TCDS.2021.3071170
- Luo, Y., and Lu, B.-L. (2021). "Wasserstein-distance-based multi-source adversarial domain adaptation for emotion recognition and vigilance estimation," in *International Conference on Bioinformatics and Biomedicine* (Los Alamitos, CA: IEEE), 1424–1428.
- Luo, Y., Zhang, S.-Y., Zheng, W.-L., and Lu, B.-L. (2018). "WGAN domain adaptation for EEG-based emotion recognition," in *International Conference on Neural Information Processing* (Cham: Springer), 275–286.
- Ma, B.-Q., Li, H., Zheng, W.-L., and Lu, B.-L. (2019). "Reducing the subject variability of EEG signals with adversarial domain generalization," in *International Conference on Neural Information Processing* (Cham: Springer), 30–42.
- Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain Comp. Interf.* 1, 66–84. doi: 10.1080/2326263X.2014.912881
- Nath, D., Singh, M., Sethia, D., Kalra, D., and Indu, S. (2020). "A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network," in *The 4th International Conference on Compute and Data Analysis* (Los Alamitos, CA), 142–147.
- Pan, B., and Zheng, W. (2021). Emotion recognition based on EEG using generative adversarial nets and convolutional neural network. *Comput. Math. Methods Med.* 2021:2520394. doi: 10.1155/2021/2520394
- Picard, R. W. (2001). "Building hal: computers that sense, recognize, and respond to human emotion," in *Human Vision and Electronic Imaging VI*, Vol. 4299 (San Jose, CA: IEEE), 518–523.
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.* 16:051001. doi: 10.1088/1741-2552/ab260c
- Rozgic, V., Vazquez-Reina, A., Crystal, M., Srivastava, A., Tan, V., and Berka, C. (2014). "Multi-modal prediction of ptsd and stress indicators," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Los Alamitos, CA), 3636–3640.
- Samek, W., Meinecke, F. C., and Müller, K.-R. (2013). Transferring subspaces between subjects in brain-computer interfacing. *IEEE Transact. Biomed. Eng.* 60, 2289–2298. doi: 10.1109/TBME.2013.2253608
- Schuller, B. W. (2018). Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends. *Commun. ACM* 61, 90–99. doi: 10.1145/3129340
- Song, Y., Zheng, Q., Liu, B., and Gao, X. (2023). EEG Conformer: convolutional transformer for EEG decoding and visualization. *IEEE Transact. Neural Syst. Rehabil. Eng.* 31, 710–719. doi: 10.1109/TNSRE.2022.3230250
- Wang, J., and Chen, Y. (2021). *Introduction to Transfer Learning*. Los Alamitos, CA: Publishing House of Electronics Industry.
- Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., et al. (2023). Generalizing to unseen domains: a survey on domain generalization. *IEEE Trans. Knowl. Data Eng.* 35, 8052–8072. doi: 10.1109/TKDE.2022.3178128
- Wu, D., Lawhern, V. J., Gordon, S., Lance, B. J., and Lin, C.-T. (2016). Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR). *IEEE Transact. Fuzzy Syst.* 25, 1522–1535. doi: 10.1109/TFUZZ.2016.2633379

## 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.

- Xu, R., Chen, Z., Zuo, W., Yan, J., and Lin, L. (2018). "Deep cocktail network: multi-source unsupervised domain adaptation with category shift," in *IEEE Conference on Computer Vision and Pattern Recognition* (Piscataway, NJ), 3964–3973.
- Yang, H., Rong, P., and Sun, G. (2021). "Subject-independent emotion recognition based on entropy of EEG signals," in *The 33rd Chinese Control and Decision Conference* (Los Alamitos, CA: IEEE), 1513–1518.
- Yang, Y., Wu, Q. J., Zheng, W.-L., and Lu, B.-L. (2017). EEG-based emotion recognition using hierarchical network with subnetwork nodes. *IEEE Transact. Cogn. Dev. Syst.* 10, 408–419. doi: 10.1109/TCDS.2017.2685338
- Zhang, Z., Zhong, S.-H., and Liu, Y. (2022). Ganser: a self-supervised data augmentation framework for EEG-based emotion recognition. *IEEE Transact. Affect. Comp.* 2048–2063, 1–17. doi: 10.1109/TAFFC.2022.3170369
- Zhao, H., Zhang, S., Wu, G., Moura, J. M. F., Costeira, J. P., and Gordon, G. J. (2018). Adversarial multiple source domain adaptation. *Adv. Neural Inf. Process. Syst.* 31, 1–12. doi: 10.1109/TPAMI.2022.3195549
- Zhao, L.-M., Yan, X., and Lu, B.-L. (2021). "Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition," in *The 35th AAAI Conference on Artificial Intelligence* (Palo Alto, CA), 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
- Zheng, W.-L., and Lu, B.-L. (2016). "Personalizing EEG-based affective models with transfer learning," in *The Twenty-Fifth International Joint Conference on Artificial Intelligence* (San Mateo, CA), 2732–2738.
- Zhou, K., Liu, Z., Qiao, Y., Xiang, T., and Loy, C. C. (2023). Domain generalization: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 4396–4415.
- Zhu, L., Ding, W., Zhu, J., Xu, P., Liu, Y., Yan, M., et al. (2022). Multisource wasserstein adaptation coding network for EEG emotion recognition. *Biomed. Signal Process. Control* 76:103687. doi: 10.1016/j.bspc.2022.103687
- Zhu, Y., Zhuang, F., and Wang, D. (2019). Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. *AAAI Conf. Artif. Intell.* 33, 5989–5996. doi: 10.1609/aaai.v33i01.33015989