

# Label-refined EEG emotion classification based on intelligent media communication

Zitian Fu<sup>1,2</sup>, Qiao Sun<sup>1,3,\*</sup>

<sup>1</sup>School of humanities, Central South University, Changsha Hunan, China

<sup>2</sup>fuzt199905@163.com

<sup>3</sup>sunqiao@csu.edu.cn

\*corresponding author

**Abstract.** With the development of cognitive computing and intelligent media communication, recognizing the user's emotion using electroencephalography (EEG) has garnered increasing attention. However, building a cross-subject emotion recognition model with good generalization performance is really difficult. To achieve remarkable performance, a label-refined domain adversarial neural network is designed and EEG samples are subdivided into 6 groups according to the differences in users' brain activities under different movie stimuli. Based on the experiment of SEED and SEED-IV datasets, the classification accuracy of our method reaches  $93.35 \pm 4.14\%$  and  $84.08 \pm 6.01\%$ , achieving the state-of-the-art performance. Experiments show that using different movie stimuli to refine labels can achieve efficient cross-subject emotion recognition.

**Keywords:** EEG emotion recognition, label refinement, domain adversarial neural network.

## 1. Introduction

The development of cognitive computing and intelligent media communication has put forward higher requirements for the recognition of cognitive states, among which recognition of emotion and understanding of users' emotional states are crucial [1]. Emotion recognition technology based on EEG signals has become a focus of research due to its low cost, anti-forgery, and anti-coercion characteristics [2-3]. EEG-based emotion recognition can help media precisely understand users' emotional responses when watching the communication content and adjust the presentation and selection of content according to users' emotional states, thereby meeting users' needs more accurately.

However, EEG signals have inter-individual differences and non-stationary characteristics, which do not conform to the assumption of independent and identically distributed training and testing data required by traditional machine learning algorithms [4]. Therefore, cross-subject EEG emotion recognition has been one of the challenges in the field of emotion recognition. To address this challenge, cross-subject emotion recognition algorithms based on transfer learning have become the mainstream of research in recent years [5-7]. These algorithms refer to the data of known subjects as a source domain and the data of unknown subjects as a target domain. The core idea is to reduce the differences between the source domain and the target domain, transfer the knowledge learned in the source domain to the target domain, and thus improve the generalization ability of the model. Unlike methods that only use source domain data to achieve task-related feature extraction, domain adaptation algorithms not only

use source domain data but also obtain data distribution from unlabeled target domain data to achieve knowledge transfer [8]. Although collecting target domain data may increase training time, domain adaptation methods have significant accuracy advantages because they obtain the data distribution of the target domain and have become an important algorithm in EEG emotion recognition transfer learning.

Traditional domain adaptation algorithms usually treat target domain data as a whole category when processing them, and ignore the possible distribution differences between data with different labels, which may lead to misleading knowledge transfer. To address this problem, few-shot domain adaptation techniques have emerged in recent years, attempting to use a small amount of labeled target domain data to transfer source domain knowledge for different label types separately in the domain adaptation process, and improving the accuracy of domain adaptation algorithms [9-11]. Some researchers have combined label information and domain types, dividing the source domain and target domain data participating in training into multiple types, further enhancing the targeting of the knowledge transfer process, and achieving advanced results.

However, research has shown that even emotional data with the same label may have large distribution differences. In contrast, different subjects may exhibit more similar cognitive pattern responses under the same movie clip stimuli [12]. This suggests that considering distinguishing different contexts with the same label and making more refined divisions of the target domain can improve the targeting of the domain adaptation process and the accuracy of cross-subject EEG emotion recognition. Therefore, scientifically dividing the target domain into subclasses to ensure the consistency of the data distribution of each subclass is crucial.

Based on the above ideas, this paper proposes a label-refined domain adversarial neural network (LR-DANN) for EEG emotion recognition, targeting different movie clips. According to the similarity and difference of the stimulus movie clips and the domain, the EEG samples participating in the training are subdivided into six types. In the domain adaptation process, the algorithm needs to distinguish movie clips while confusing domain information, thereby achieving more efficient cross-domain emotion representation. Experiments show that label type refinement can effectively improve the targeting of the domain adaptation process.

## 2. Related works

### 2.1. EEG emotion recognition

Cognitive science research indicates a close relationship between brain electrical activity and emotions and cognition [13]. Based on EEG signals, researchers can extract emotion-related information to establish computational models of emotions, enabling the detection and recognition of different emotional states. Researchers can apply signal processing techniques and deep learning algorithms to analyze EEG signals, extract features related to emotions and cognition, thereby achieving detection of internal states in humans.

As a method of audiovisual stimulation, movies and other video stimuli are widely used in emotion recognition research. Similar to real-life scenarios, movies and videos are rich in content, dynamically plotted, and highly realistic, making them ecologically valid stimuli. Consequently, an increasing number of emotion recognition studies are incorporating audiovisual stimuli to induce emotions. SEED [14] and SEED-IV [15] are publicly available EEG databases built upon audiovisual stimuli, widely utilized as foundations in many emotion recognition studies.

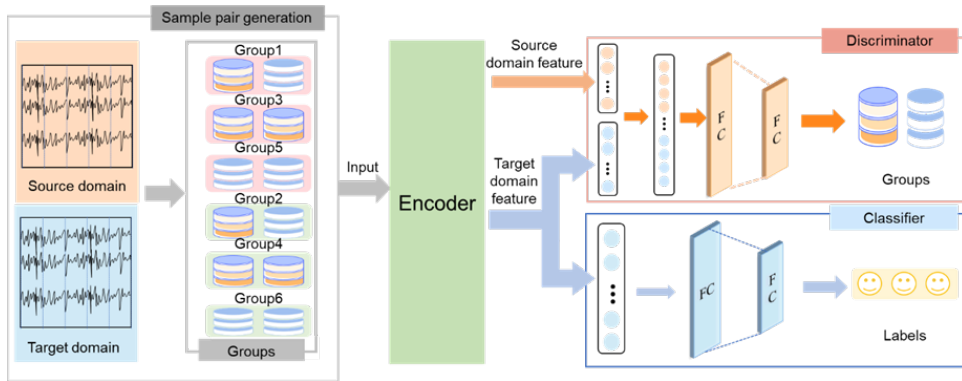
### 2.2. Domain adaption emotion recognition based on EEG

Due to cost and efficiency issues in data collection, it is impractical to gather large-scale data from every user for model training in brain-machine interface systems based on physiological signals. Consequently, the impact of individual differences cannot be entirely avoided. When significant individual differences exist in participants' EEG signals, traditional machine learning algorithms often struggle to construct accurate cross-subject emotion recognition models. To address this challenge, the prevailing approach involves using domain adaptation algorithms to develop more generalized emotion recognition models.

The goal of these algorithms is to learn common attributes between the source and target domains and transfer knowledge from the source domain to tasks in the target domain.

For instance, Zheng et al. proposed an emotion recognition model based on traditional domain adaptation algorithms, achieving a 20% accuracy improvement in cross-subject emotion recognition [16]. Wang et al. also proposed a few-label adversarial domain adaption (FLADA) method for cross-subject EEG emotion recognition tasks. This method combines adversarial domain adaptation with deep adaptive transformations, achieving robust emotion recognition across subjects with small-scale EEG datasets containing few labels [9]. Drawing inspiration from domain adaption neural networks, this study refines emotion labels to handle domain ambiguity while performing emotion classification tasks on movie clips. This approach enables more efficient cross-domain emotion representation.

### 3. Method

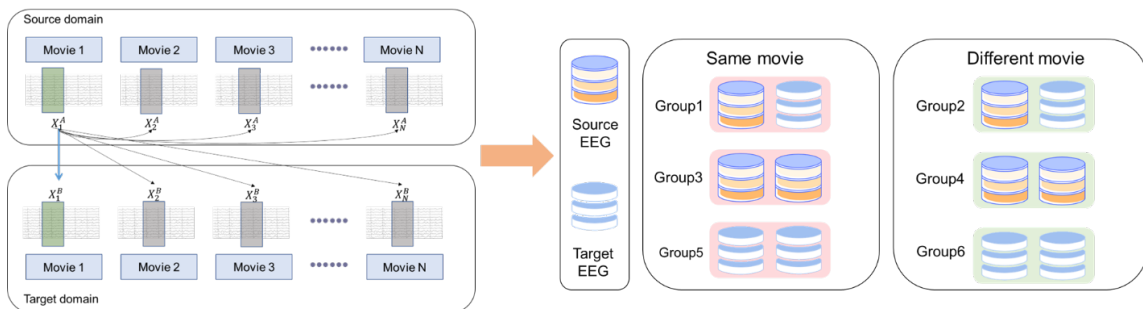


**Figure 1.** An overview of the proposed LR-DANN.

The algorithm in this paper is divided into two stages. In the first stage, all source domains are adapted to the target domain one by one according to the structure shown in Figure 1. In the second stage, the top K performing source domains are integrated and classified to achieve cross-subject emotion recognition. The first stage comprises four modules: sample pair generation, encoder, domain discriminator, and movie classifier. The remaining part of this section will introduce each module sequentially.

#### 3.1. Sample pair generation

To make the domain adaptation process more targeted, the algorithm in this paper refines the labeling of the training data by categorizing it into N movie stimuli based on movie segments. For instance, the SEED dataset has 3 emotional categories, each elicited by 5 different stimulus videos, resulting in 15 distinct movie stimulus labels. Subsequently, the training samples are categorized into 6 types based on the movie stimuli they belong to and the domain types they are part of. The categorization process is illustrated in Figure 2.



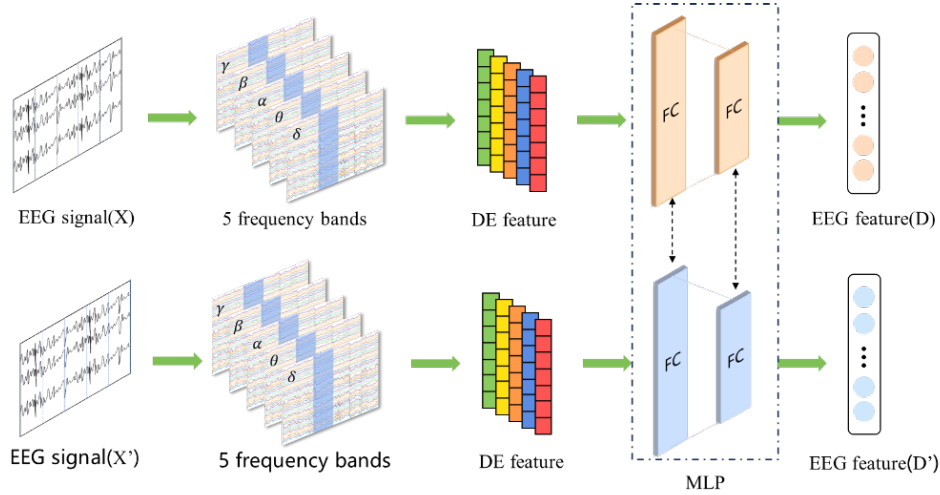
**Figure 2.** Type division of EEG samples.

This paper divides the original EEG data into samples of 1-second duration each, and randomly selects pairs of samples from all EEG samples to form sample pairs  $(X, X')$ . To distinguish the domain source of samples using superscripts: samples from the source domain are denoted as  $X^A$ , samples from the target domain are denoted as  $X^B$ ; To differentiate samples based on movie stimuli using subscripts: EEG samples from the  $n$ -th movie stimulus are denoted as  $X_n$ . The sample pairs can be categorized into 6 groups based on the type of movie stimulus and their domain of origin, as shown in Figure 2.

### 3.2. Encoder

To obtain emotion-related EEG features, we compute the differential entropy (DE) across the segmented sample pairs  $(X, X')$  as outlined in Section 3.1. These features are utilized for cross-subject domain-adaptive learning. We implement the methodology proposed by Shi et al. for this purpose [17]. For each EEG sample, we initially decompose the training dataset using Butterworth filters into 5 frequency bands ( $\delta$ [0.5~4 Hz],  $\theta$ [4~8 Hz],  $\alpha$ [8~14 Hz],  $\beta$ [14~31 Hz] and  $\gamma$ [31~51 Hz]). Next, we extract DE features from the 5 frequency bands individually, arranging them in a one-dimensional vector in channel order from top to bottom. This results in five feature vectors. Subsequently, we concatenate these feature vectors from lower to higher frequency bands to form a one-dimensional DE feature vector denoted as  $E$ . We apply these operations sequentially to each EEG sample from the sample pairs  $(X, X')$  obtained in Section 3.1 and get DE pairs  $(E, E')$ .

We use a multi-layer perceptron (MLP) with shared parameters to process the extracted DE pairs  $(E, E')$ . Each MLP consists of fully connected hidden layers and an output layer. The MLP functions as a feature extractor, where the hidden layers act to map the raw data into a high-dimensional feature space. During training, parameter sharing is implemented between the two MLPs. Each neuron computes a weighted set of inputs, passes them through a ReLU activation function to produce nonlinear outputs. The final output is represented as  $(D, D')$ , as depicted in Figure 3.



**Figure 3.** The process of EEG feature representation.

### 3.3. Domain discriminator

To confound the source and target domains and facilitate cross-subject emotional knowledge transfer, we designed a domain discrimination model. This model categorizes sample pairs based on EEG features  $(D, D')$  into 6 types (denoted as G1 to G6). We implemented a multi-class domain-category discriminator (DCD) for this purpose. The structure of the DCD includes two fully connected layers with a softmax activation function, trained using standard cross-entropy loss as shown in Equation (1).

$$Loss(P) = -E\left[\sum_{k=1}^6 (y_{g_k} \log(P(G_k)))\right] \quad (1)$$

In the context provided,  $G_k$  represents the type of sample pair,  $y_{g_k}$  denotes the label for  $G_k$ , and  $P$  represents the multi-class domain-category discriminator (DCD). We continually update the MLP to confuse the DCD, making it unable to distinguish between source and target domain data. Specifically, the DCD should not distinguish between sample pair types 1, 3, and 5, nor should it distinguish between types 2, 4, and 6. Therefore, the DCD loss function can be transformed into Equation (2).

$$\begin{aligned} Loss(P) = & -E[y_{G_3} \log(P(G_3)) + y_{G_3} \log(P(G_1)) \\ & + y_{G_4} \log(P(G_2)) + y_{G_4} \log(P(G_6))] \end{aligned} \quad (2)$$

### 3.4. Movie classifier

While confounding the source and target domains, we train a movie stimulus classifier using the EEG features( $D, D'$ ) obtained in Section 3.2 to predict the stimuli. Training is performed using the standard cross-entropy loss function, with classification losses represented as Equations (3) and (4).

$$Loss(D) = -\sum_{i=1}^M y_i \log(C(D)) \quad (3)$$

$$Loss(D') = -\sum_{i=1}^M y_i \log(C(D')) \quad (4)$$

In this context,  $y_i$  represents the genre tag for thriller movies,  $C$  denotes the thriller movie classifier, and  $M$  stands for the number of segments in the movie.

The joint adversarial training of the thriller movie classifier and multi-class domain discriminator involves specific steps. In the first step, freeze the encoder and the movie classifier, and train domain discriminator on pairs of samples from 6 categories. In the second step, freeze domain discriminator and train the encoder and movie classifier while maintaining high accuracy in classifying movies. Repeat these steps until the loss converges to a preset value. The total training loss is shown in formula (5).

$$Loss = Loss(D) + Loss(D') - Loss(P) \quad (5)$$

### 3.5. Ensemble classification

Fixing the target domain subjects, according to the domain adaptation steps described in sections 3.1 to 3.4, adapt each of the  $P$  source domains individually to the target domain. Each source domain is trained with the target domain to obtain one domain adaptation model (as shown in Figure 1). In the second stage, we evaluate the  $P$  domain adaptation models obtained in the first stage using a few labeled target domain data. We select the top  $K$  models with the best prediction results to classify the target domain EEG data for emotion classification.

We aggregate all predicted results using equal weights, as shown in Equation (6).

$$pred = \text{sigmoid}\left(\sum_{i=1}^K (w \times R_i)\right) \quad (6)$$

Here,  $w$  represents the equal weights of domain training models.  $P$  is the number of domain adaptation models and  $K$  is the number of selected domain adaptation models.  $R_i$  denotes the emotion prediction result of the  $i$ -th domain adaptation model.

## 4. Experiment

### 4.1. Dataset

This study utilizes the publicly available SEED dataset, collected by Shanghai Jiao Tong University for EEG-based emotion recognition. The dataset includes data from 15 participants (7 males, 8 females). There are 15 movie stimuli, each consisting of a 4-minute segment from different movies, inducing 3

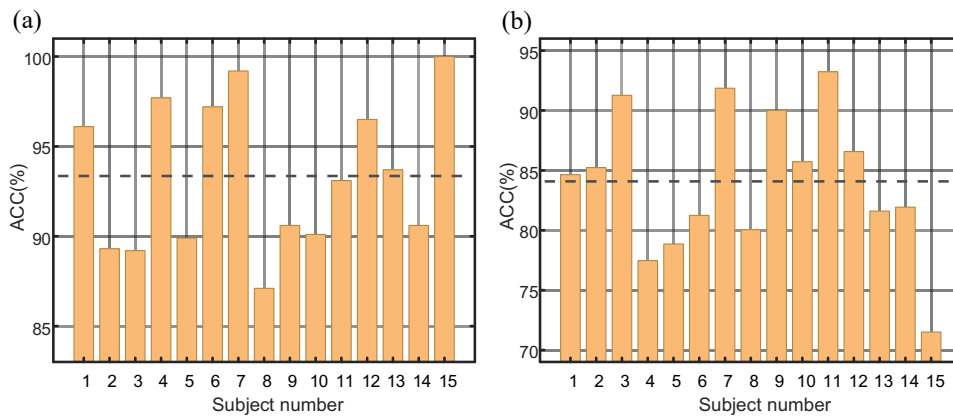
emotional categories (positive, neutral, and negative), with 5 stimuli per emotion category. The experiment is divided into four stages: a 5-second pre-stimulus phase, a 4-minute movie stimulus phase, a 45-second self-assessment phase, and a 20-second rest phase.

The SEED-IV dataset provides EEG emotion data in 4 categories: neutral, sad, fear, and positive. Similar to the SEED dataset, it includes data from 15 subjects. Each subject undergoes 3 experiments, divided into four stages: a 5-second pre-stimulus phase, approximately 2 minutes of movie stimulus phase, a 45-second self-assessment phase, and a 20-second rest phase. There are 24 movie stimuli per experiment, with 6 stimuli corresponding to each emotional category.

## 4.2. Results

**4.2.1. Cross-subject emotion recognition performance.** To evaluate the algorithm in this study, each participant randomly selected one out of three experimental groups as their test data. We employed leave-one-subject-out cross-validation to train and test the EEG data of 15 subjects from the SEED and SEED-IV datasets. In each experiment, one subject was designated as the target domain, while the remaining 14 subjects served as the source domain.

In this experiment, the target domain utilized very limited labeled data to classify the remaining unlabeled data. For the SEED dataset, 18 randomly selected labeled samples per movie stimulus were used as calibration data, resulting in  $L=90$  calibration samples per emotion category. The average accuracy (ACC) and standard deviation (STD) for emotion classification across all participants were 93.35% and 4.14%, respectively, as illustrated in Figure 4 (a). On the SEED-IV dataset, calibration data was selected in the same proportion as the SEED dataset. The model achieved an average ACC of 84.08% and a STD of 6.01%, as shown in Figure 4 (b).



**Figure 4.** The performance of cross-subject emotion recognition on SEED dataset and SEED-IV dataset

**Table 1.** The comparison of data-independent results between the proposed algorithm and others on SEED dataset and SEED-IV dataset

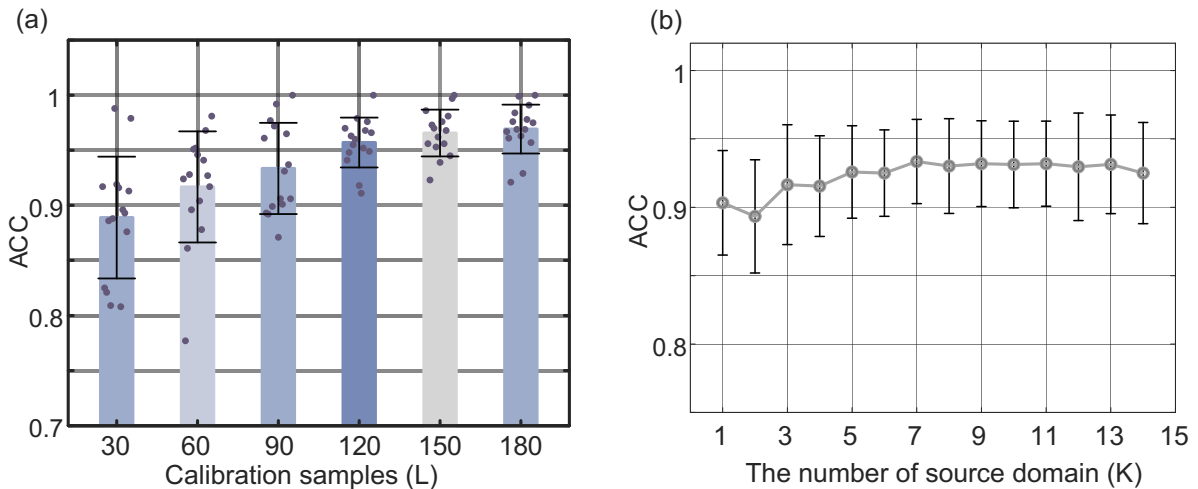
Model	SEED		SEED-IV	
	ACC (%)	STD (%)	ACC (%)	STD (%)
Source-only	60.27	18.89	40.50	10.05
DANN	75.08	11.18	54.63	8.03
R2G-STNN	84.16	7.63	-	-
MTL	88.92	10.35	-	-
MS-MDA	80.62	11.03	57.92	10.12
TANN	-	-	68.00	8.35
GMSS	76.04	11.91	62.13	11.91
FLADA	89.32	0.86	82.91	8.35
<b>LR-DANN (ours)</b>	<b>93.35</b>	<b>4.14</b>	<b>84.08</b>	<b>6.01</b>



To further validate the performance of the proposed algorithm in cross-subject emotion recognition, this study compares LR-DANN with several state-of-the-art research algorithms (DANN [18], R2G-STNN [19], MTL [11], FLADA [9], MS-MDA [20], TANN [21], and GMSS [22]). All algorithms were evaluated using the same parameters, and the results are presented in Table 1. It is evident from the table that the proposed algorithm achieves significantly higher accuracy compared to the other algorithms tested.

**4.2.2. Model evaluation based on calibration data.** In transfer learning, the quantity of calibration data ( $L$ ) in the target domain is an important parameter. Generally, a larger amount of calibration data in the target domain leads to better model performance. However, obtaining a substantial amount of calibration data can be challenging in practical applications. Therefore, it is necessary to assess the algorithm's performance under different values of  $L$ . This study analyzed the SEED dataset where a total calibration data quantity  $L$  increasing from 30 to 180. The model performance remains consistently high across different values of  $L$ , as shown in Figure 5 (a).

**4.2.3. Model evaluation based on the selection number of source domains.** In multi-source domain adaptation, the number  $K$  of participating source domains is another critical parameter. When choosing multiple source domains, the model can benefit from a richer diversity of domain-specific information. We analyzed the experimental results of LR-DANN under different numbers  $K$  of source domains, as shown in Figure 5 (b). As the number of source domains increases, the overall accuracy of the model predictions also increases. However, once the number of source domains exceeds 7, the model performance maintains an accuracy level of around 93%.



**Figure 5.** Performance comparison of LR-DANN with different calibration samples and source numbers.

## 5. Conclusion

In the process of multi-source domain adaptation in the target domain, fine-grained labeling can effectively enhance the specificity of domain adaptation. This study uses the variability of EEG responses under different movie stimuli to refine emotional labels, categorizing emotions of different subjects into intra-group divisions based on the specific movie stimuli. This approach fully utilizes the individual differences in brain activity patterns activated by different emotional stimuli. Moreover, the study employs ensemble classification from multiple source domains, making it easier to identify optimal domain parameters for the EEG data under examination, thereby achieving efficient cross-subject EEG emotion recognition.

## References

- [1] Cowie R, Douglas-Cowie E, Tsapatsoulis N, et al. (2001). Emotion recognition in human-computer interaction[J]. *IEEE Signal processing magazine*. 18(1): 32-80.
- [2] Alarcão S M, Fonseca M J. (2019). Emotions Recognition Using EEG Signals: A Survey[J]. *IEEE Transactions on Affective Computing*. 10(3): 374-393.
- [3] Li G, Chen N, Jin J. (2022). Semi-supervised EEG emotion recognition model based on enhanced graph fusion and GCN[J]. *Journal of Neural Engineering*. (No.2): 26039.
- [4] Peng Y, Liu H, Kong W, et al. (2022). Joint EEG Feature Transfer and Semi-supervised Cross-subject Emotion Recognition[J]. *IEEE Transactions on Industrial Informatics*. 1-12.
- [5] Wan Z, Yang R, Huang M, et al. (2021). A review on transfer learning in EEG signal analysis[J]. *Neurocomputing*. 1-14.
- [6] Ma Y, Zhao W, Meng M, et al. (2023). Cross-Subject Emotion Recognition Based on Domain Similarity of EEG Signal Transfer Learning[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 936-943.
- [7] Quan J, Li Y, Wang L, et al. (2023). EEG-based cross-subject emotion recognition using multi-source domain transfer learning[J]. *Biomedical Signal Processing and Control*. 104741.
- [8] Zhou K, Liu Z, Qiao Y, et al. (2023). Domain Generalization: A Survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. (No.4): 4396-4415.
- [9] Wang Y, Liu J, Ruan Q, et al. (2021). Cross-subject EEG emotion classification based on few-label adversarial domain adaption[J]. *Expert Systems with Applications*.
- [10] Fan J, Wang Y, Guan H, et al. (2022). Toward few-shot domain adaptation with perturbation-invariant representation and transferable prototypes[J]. *Frontiers of Computer Science*. (No.3): 163347.
- [11] Li J, Qiu S, Shen Y, et al. (2020). Multisource Transfer Learning for Cross-Subject EEG Emotion Recognition[J]. *IEEE transactions on cybernetics*. (No.7): 3281-3293.
- [12] Shen X, Liu X, Hu X, et al. (2022). Contrastive Learning of Subject-Invariant EEG Representations for Cross-Subject Emotion Recognition[J]. *IEEE Transactions on Affective Computing*. (No.99): 1.
- [13] Lin Y P, Wang C H, Jung T P, et al. (2010). EEG-based emotion recognition in music listening[J]. *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp.1798-1806.
- [14] Duan R-N, Zhu, J-Y and Lu B-L. (2013). Differential entropy feature for eeg-based emotion classification. In 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, pp. 81-84.
- [15] Zheng W-L, Liu W, Lu Y, Lu B-L, and A. Cichocki. (2018). Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1110-1122.
- [16] Zheng W-L, Lu B-L. (2016). Personalizing EEG-based affective models with transfer learning[C]//*Proceedings of the 25th International Joint Conference on Artificial Intelligence*. 2732-2738.
- [17] Shi L C, Jiao Y Y, Lu B L. (2013). Differential entropy feature for EEG-based vigilance estimation[C]. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, pp. 6627-6630.
- [18] Ganin Y, Ustinova E, Ajakan H. (2016). Domain-Adversarial Training of Neural Networks[J]. *Journal of Machine Learning Research*.
- [19] Li Y, Zheng W, Wang L, et al. (2019). From Regional to Global Brain: A Novel Hierarchical Spatial-Temporal Neural Network Model for EEG Emotion Recognition[J]. *IEEE Transactions on Affective Computing*.
- [20] Chen H, Jin M, Li Z, Fan C, et al. (2021). MS-MDA: Multisource Marginal Distribution Adaptation for Cross-Subject and Cross-Session EEG Emotion Recognition[J]. *Frontiers in neuroscience*, Vol.15: 778488.



- [21] Li Y, Fu B, Li F, et al. (2021). A novel transferability attention neural network model for EEG emotion recognition[J]. *Neurocomputing*, Vol.447: 92-101
- [22] Li Y, Chen J, Li F, et al. (2023). GMSS: Graph-Based Multi-Task Self-Supervised Learning for EEG Emotion Recognition[J]. *IEEE Transactions on Affective Computing*, Vol.14(3): 2512-2525