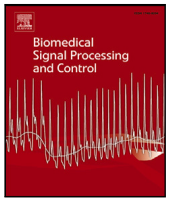
Biomedical Signal Processing and Control 97 (2024) 106716

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



A novel multi-source contrastive learning approach for robust cross-subject emotion recognition in EEG data

Xin Deng [a], Chenhui Li [a], Xinyi Hong [a], Huaxiang Huo [a], Hongxing Qin [b,*]

[a] The Key Laboratory of Data Engineering and Visual Computing, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[b] College of Computer Science, Chongqing University, Chongqing, 400044, China

ARTICLE INFO

Keywords:
Brain–computer interface
Electroencephalograph(EEG)
Emotion recognition
Domain adaptation
Contrastive learning

ABSTRACT

Emotion Brain Computer Interface (BCI) based on Electroencephalography (EEG) is a significant branch in the field of affective computing. However, the variability in emotional feedback among subjects facing the same emotional stimuli and the potential presence of noisy labels in the data significantly limit the effectiveness and generalizability of EEG-based emotion recognition models. In this paper, we propose the **M**ulti-**S**ource **C**ontrastive **L**earning Transfer Learning Model (MSCL) to address the emotion recognition in cross-subject scenarios. The MSCL model consists of three components: Firstly, the unsupervised and supervised contrastive learning are utilized to learn the differences and commonalities among different individuals. Secondly, domain adaptation methods and feature learning in multiple domains enhance the interaction of information between source and target domains, thereby improving the model's generalization across different individuals. Thirdly, in order to mitigate the detrimental effects of noisy labels on the model, this research dynamically allocate the weights to source domains similar to the target domain based on behavioral characteristics within the target domain and employ the corresponding noise learning methods. In our experiments, MSCL achieves excellent cross-subject emotion recognition performance on the CEED Dataset and the SEED Dataset. Our research highlights the significant capability of MSCL model in addressing the issues associated with inter-subject variability and label noise within EEG-based emotion BCI systems.

1. Introduction

Emotions significantly impact daily human experiences, shaping decision-making, social interactions, intellectual processes, and perceptions [1]. The recognition of emotions is fundamental to the development of efficient affective BCIs, which seek to equip machines with the capability to accurately identify human emotional states. Typically, the emotion assessment incorporates four data modalities: facial expressions, vocal patterns, textual analysis, and physiological signals [2]. In recent years, there has been a marked increase in interest towards EEG-based emotion recognition within the fields of affective computing and human emotion analysis, highlighting its importance in advancing the understanding and interaction between humans and machines [3]. Unlike behavioral techniques that record facial expressions, body postures, and sounds, EEG provides a more direct and objective measurement of human emotional responses, which cannot be easily masked or consciously suppressed. Compared to other neuroimaging techniques like fMRI and MEG, EEG is favored in practical applications for its portability and cost-effectiveness [4,5].

Extracting the effective features is crucial for efficiently deriving meaningful information from EEG signals for emotion recognition tasks. Power Spectral Density (PSD) is a significant feature in emotion recognition research, transforming raw signals into a power spectrum that varies with frequency, allowing for a clear and intuitive observation of the signal's frequency components [6]. The most common method for power spectrum estimation is classical spectral estimation, which is achieved through Fourier Transform. Differential Entropy (DE) features [7], equivalent to the logarithmic energy spectrum in specific frequency bands, have been widely adopted in state-of-the-art emotion recognition approaches. Studies indicate that DE features in the Beta and Gamma bands in the temporal region are closely associated with emotions [8]. Additionally, researchers have explored the use of the Tunable Q-factor Wavelet Transform (TQWT) [9] for feature extraction in emotion recognition tasks. Furthermore, a novel feature selection method based on greedy algorithm and Max-Relevance and Min-Redundancy(Greedy-mRMR) [10] has been proposed, employing a greedy algorithm to acquire top-ranking features and using the mRMR algorithm to select the most relevant and minimally redundant features.

In addition to extracting effective features for EEG-based emotion recognition, researchers have attempted to use deep neural networks

* Corresponding author.
E-mail address: qinhx@cqu.edu.cn (H. Qin).

https://doi.org/10.1016/j.bspc.2024.106716
Received 5 March 2024; Received in revised form 15 June 2024; Accepted 1 August 2024
Available online 9 August 2024

to enhance model performance while avoiding manual feature extraction. The implicit correlations among different channels are crucial indicators for emotion recognition. Due to their suitability for processing two-dimensional data and extracting joint information between channels, convolutional neural networks (CNNs) have been applied to emotion recognition tasks [11,12]. Researchers have also explored spatial and temporal relationships between different EEG channels using graph neural networks (GNNs) [13] and long short-term memory networks (LSTM) [9]. Furthermore, the capsule networks [14] and attention mechanisms [15] have also been applied to learn emotion-related representations. These methods leverage deep neural networks to enhance the model's representation capability and avoid manual feature extraction.

Emotion experiments are time-consuming and labor-intensive. To obtain satisfactory emotion recognition performance, researchers need to collect a significant amount of EEG data from subjects, with data collection for a single participant typically requiring more than an hour [16]. This labor-intensive and time-consuming training process is a major bottleneck for the practical application of EEG-based emotion recognition. Consequently, it is desirable to develop methods that exhibit strong cross-subject generalizability, particularly for accommodating new users. Nonetheless, models that are trained on EEG data from one or several subjects frequently encounter challenges in generalizing to new individuals. This limitation is primarily attributed to the variability in emotional responses to stimuli across different subjects, as well as the financial and temporal constraints associated with conducting emotion collection experiments [17].

Furthermore, when individuals are exposed to emotional stimuli, such as emotional videos, the resulting physiological responses can vary significantly between people. This variability can lead to noisy labels in the EEG signals collected during these labor-intensive emotional experiments. These noisy labels can negatively impact the training of emotion recognition models. To address this issue, researchers have explored the use of the capsule network with a joint optimization strategy (JO-CapsNet) [17]. This method updates network parameters based on the loss function of the capsule network and updates pseudo-labels by predicting the presence probability of class labels from the capsule network's output. This alternating update strategy has been shown to be mutually beneficial in correcting noisy labels. Additionally, researchers have proposed a novel transfer learning framework based on prototype representation and pairwise learning (PR-PL) [18]. This framework aims to learn discriminative and generalized prototype representations of emotional revelations among individuals. By formulating emotion recognition as a pairwise learning task, this approach reduces the dependence on precise label information.

To overcome the aforementioned issues, we propose the MSCL. MSCL hypothesizes that when subjects receive the same category of emotional stimuli (emotional videos in our study), their neural activities are in similar states. Based on this, we use the contrastive learning to learn detailed neural activity representations, obtaining invariant and specific attributes from different subjects' EEG signals. Specifically, our MSCL model comprises three stages: a common space stage, a subject-specific subspace stage, and a classification stage. In the common space stage, we use a shared feature extractor to map the source domain data and the target domain data into a common space and use the contrastive learning to obtain invariant features of the data and control their sample distribution. In the subject-specific subspace stage, we pair-wise map the source domain features and the target domain features into different subject-specific feature subspaces and use the contrastive learning to obtain the subject-specific features extracted from the subspaces. In the prediction phase, classifiers that have been trained on various source domains are assigned weights dynamically, according to the performance of the target domain data. This method facilitates the recognition of emotions based on the features specific to the target domain, enhancing the adaptability and effectiveness of the emotion recognition process.

In this study, we have made the following contributions:

- We propose the MSCL, which can effectively avoid the variability among subjects and enhances the accuracy of cross-subject emotion recognition.
- We employ contrastive learning in both the common space and the subject-specific subspace. This approach is effective in learning both general response features and individual-specific characteristics, thereby obtaining comprehensive emotional information.
- We integrate a dynamic selection process for pinpointing the most pertinent source domain data. Moreover, we implement noise learning strategies to alleviate the impact of noisy labels, improving the reliability and precision of our emotion recognition endeavors.
- We extensively evaluate our model on the SEED Dataset and the CEED Dataset. The results show that our method can outperforms the traditional transfer learning and state-of-the-art domain adaptation methods.

2. Related work

In this chapter, we review cross-subject emotion recognition and contrastive learning.

2.1. Cross-subject emotion recognition

Cross-subject emotion recognition based on EEG is an important research direction in the field of emotion recognition in recent years [8]. Due to physiological and psychological differences between subjects, cross-subject emotion recognition faces significant challenges, particularly in effectively handling individual differences and improving model generalization. To overcome these issues, researchers have employed various machine learning and deep learning techniques, such as transfer learning and domain adaptation, to optimize and adjust models for better transferability and adaptability across different individuals. Researchers have applied Domain Adaptation (DA) methods to cross-subject emotion recognition. Typical DA algorithms aim to learn subject features by reducing data distribution differences between source and target domains, enabling models trained on source data to make predictions for target domain data. Existing research indicates that DA is an effective method for emotion recognition. However, most DA methods only minimize the marginal distribution differences between source and target domains. For instance, the Deep Adaptation Network (DAN) and the Joint Adaptation Network (JAN) minimize the Maximum Mean Discrepancy (MMD) and Joint MMD (JMMD), respectively. These measures only consider domain-level differences, which are too coarse to learn distinguishable and generalizable features. Dynamic Domain Adaptation (DDA) algorithms have been proposed, but they often overlook individual differences while focusing on class information between source and target domains. Multi-source domain adaptation methods, such as MS-MDA and MSMRA, have shown promising results on the SEED Dataset. Adversarial Discriminative Temporal Convolutional Networks (AD-TCNs) [19] has been proposed to ensure the invariance of feature map representations across different domains and bridge the gap between them.

Additionally, researchers have explored other methods to overcome the variability among subjects. For instance, the spatio-temporal self-constructing graph neural network (ST-SCGNN) [20] has been proposed, which can dynamically update the graph structure of the neural network based on input signals to address cross-subject issues. A variational instance-adaptive graph method (V-IAG) [21] has also been introduced. This method simultaneously captures individual dependencies between different EEG electrodes and estimates latent uncertain information, thereby enhancing cross-subject emotion recognition performance. Furthermore, a graph-based multi-task self-supervised learning model (GMSS) [22] has been proposed. This model integrates multiple self-supervised tasks through multi-task learning to learn more

general representations, reducing the chances of overfitting in emotion recognition tasks and addressing the issue of lacking generalization in emotion features. A combination of multi-scale residual network (MSRN) and meta-transfer learning (MTL) [23] strategies has also been proposed. The MSRN strategy learns multi-scale EEG features and interactions between different brain regions, while the MTL strategy leverages the characteristics of meta-learning and transfer learning to significantly reduce individual differences.

2.2. Contrastive learning

Contrastive learning is a popular self-supervised learning algorithm that can achieve excellent feature representations without the need for labeled data or with only a small amount of labeled data. The principle of contrastive learning lies in learning an embedding space where samples from the same class are pulled closer together, while samples from different classes are pushed apart [24]. To construct contrastive loss, positive and negative instance features are generated for each sample.

Different contrastive learning methods have varying strategies for generating instance features. Memory bank methods [25] store the features of all samples calculated in the previous step. End-to-end methods [26] use all samples in the current mini-batch to generate instance features. Momentum encoder methods [25] encode running samples through momentum-updated encoders and maintain a queue of instance features. For a given training set $X = \{x_1, x_2, \ldots, x_n\}$, it is mapped to $V = \{v_1, v_2, \ldots, v_n\}$ through an embedding function f_θ. Contrastive learning is achieved by optimizing the loss function, which is expressed as:

$$L_{con} = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\exp(v_i \cdot v_i'/\tau)}{\sum_{j=0}^{r} \exp(v_i \cdot v_j'/\tau)} \tag{1}$$

where v_i and v_i' are the sample i and its positive embedding, v_j' includes one positive embedding from other samples and r negative embeddings, and τ is the temperature hyperparameter. In MoCo [24], these embeddings are obtained by feeding x_i into a momentum encoder parameterized by θ', $v_i' = f_{\theta'}(x_i)$, where θ' is the moving average of the parameters θ.

Recently, researchers have started applying contrastive learning to physiological signals. For example, the concept of Mixup [27] from computer vision has been extended to time series analysis and combined with contrastive learning to create a new ECG signal augmentation method [28]. For EEG analysis, inspired by SimCLR [29], contrastive learning models have been proposed for sleep stage classification, clinical anomaly detection, and emotion recognition [30].

In the field of emotion recognition, a contrastive learning method for inter-subject alignment (CLISA) [31] has been proposed, significantly improving the performance of cross-subject EEG-based emotion recognition. Additionally, a two-phase prototypical contrastive domain generalization framework (PCDG) [32] has been introduced, combining a novel convolutional neural network based on residual and CBAM blocks with prototypical contrastive learning, achieving excellent cross-subject emotion recognition results. Contrastive learning has also been applied for the first time to cross-corpus emotion recognition, aligning embeddings in a shared latent time-frequency space, and achieving state-of-the-art results on two recognized datasets. Today, contrastive learning is widely applied in the analysis of physiological signals and has shown promising results in emotion recognition.

3. Methodology

In this study, we address a classification task centered on domain adaptation for EEG signals elicited by emotional stimuli. Specifically, we are confronted with two distinct data distributions: the source domain, denoted as D_S, and the target domain, referred to as D_T. The source domain, D_S, is characterized by the set $D_S = \{X_j^S, Y_j^S\}_{j=1}^{M_S}$, where M_S represents the number of subjects within the source domain. In our experimental setup, we have accessed to labeled samples original from various sources, denoted as $S = \{x_i^S, y_i^S\}_{i=1}^{N_S} \sim D_S$. Here, x_i^S corresponds to an individual sample from the source domain, accompanied by its respective label y_i^S, and N_S denotes the total number of samples within the source domain. Furthermore, we possess the unlabeled samples stemming from the target domain, denoted as $T = \{x^{T_i}\}_{i=1}^{N_T} \sim D_T$, where x_i^T signifies an unlabeled sample from the target domain. N_T represents the total number of samples in the target domain. Our primary objective is to train a model, denoted as Φ, using the source domain D_S in order to achieve superior classification performance on the unlabeled target domain D_T.

Table 1
Notation table.

Symbol	Definition
X	Instance set (matrix)
Y	Label set (matrix)
S	Source domain
T	Target domain
C	Number of labels
B	Minibatch
M^S	Number of source domain
M^T	Number of target domain
Q	Common feature
Z	Subject specific feature
ϕ, Φ	Mapping function
$\mathcal{H}$	Reproducing kernel Hilbert space
CFE	Common feature extractor
SFE	Subject-specific feature extractor
SSC	Subject-specific classifier
x	Feature vector
y	Label vector
q	Feature vector after CFE
z	Feature vector after SFE
μ	Prototype of source domain
$\hat{y}$	Pseudo label

For clarity, Table 1 is listing the symbols and their definitions used in subsequent sections. The framework of our proposed MSCL model is depicted in Fig. 1, consisting of three stages: the common space stage, the subject-specific space stage, and the classification stage. More specifically, for the input data originating from M_S distinct source domains $\{x_i^S, y_i^S\}_{i=1}^{N_S}$, and target domain data $\{x^{T_i}\}_{i=1}^{N_T}$, they are transformed into a shared feature space via the Common Feature Extractor (CFE), yielding $\{q_i^S, y_i^S\}_{i=1}^{N_S}$ and $\{q^{T_i}\}_{i=1}^{N_T}$. During the common space stage, we focus on acquiring common physiological features from all subjects when exposed to emotional video stimuli. Subsequently, each subject's features are mapped to their individualized subspace using the Subject-specific Feature Extractor (SFE), resulting in $\{Z_j^S, Y_j^S\}_{j=1}^{M_S}$ and $\{Z_j^T\}_{j=1}^{M_T}$. These subject-specific features, capturing variations among subjects in response to emotional stimuli, are leveraged in the subject-specific space stage. Lastly, the classification stage is to generate the predictions using classifiers trained within each subject-specific subspace.

3.1. Problem formulation

See Table 1.

3.2. Feature extraction

The Differential Entropy (DE) features represent the logarithmic energy spectra in specific frequency bands [7] and are widely utilized in advanced emotion recognition methods due to their capacity to identify distinct frequency band patterns [33]. In the scope of this study, we extract the DE features from five frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), and gamma (31–50 Hz). Assuming the EEG

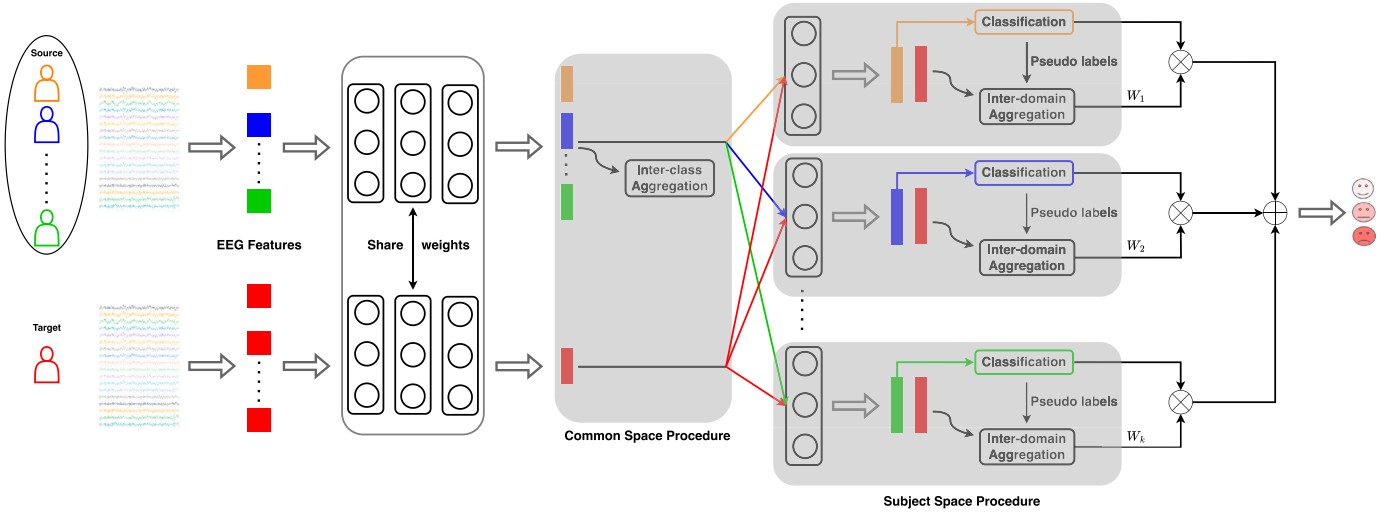


Fig. 1. The proposed model workflow for EEG-based emotion recognition. EEG features from both source and target subjects are extracted and processed through a common space procedure with shared weights, enabling inter-class aggregation. Subsequently, each domain undergoes a subject-specific procedure including classification, pseudo-label generation, and inter-domain aggregation. The outputs from each domain are then weighted and combined to produce the final emotion recognition result.

data follows a Gaussian distribution $N(\mu, \sigma^2)$, the DE feature can be defined as:

$$\begin{aligned} \mathrm{DE} &= -\int f(\mathrm{x}) \log(f(\mathrm{x})) d\mathbf{x} \\ &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}\right) \log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right. \\ &\left. \exp\left(\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}\right)\right] d\mathbf{x} = \frac{1}{2}\log 2\pi e\sigma^2 \end{aligned} \tag{2}$$

This formula effectively computes the DE feature by integrating the product of the Gaussian probability density function and its logarithm, yielding a representation of the log-energy spectrum for the specified frequency bands.

3.3. Common space procedure

In the Common Space Procedure, the goal is to map the source domain data from all subjects, along with the target domain data, into a unified shared feature subspace. Within this subspace, the common features elicited by the same emotional stimuli across subjects are extracted. To optimize the network structure and conserve computational resources, the Common Feature Extractor (CFE) is composed of a three-layer Multi-Layer Perceptron (MLP). When the multiple source domain data and the target domain data are input into the CFE, it extracts low-level domain-invariant features.

3.3.1. Inter-class aggregation

In the shared feature subspace, we employ the contrastive learning to enhance inter-class clustering. Our goal for this approach is to ensure a balanced distribution of data across domains within the shared subspace, while also capturing the universal emotional features exhibited by different subjects.

For any given source domain sample x_i^S, it is transformed to the shared feature subspace through the CFE, resulting in the feature q_i^S. The contrastive loss L_{con1} for each sample is calculated as:

$$L_{con1} = -\frac{1}{|P_1(x)|} \sum_{q^+ \in P_1(x)} \log \frac{\exp\left(q_i^S \cdot q_j^+ / \tau_1\right)}{\sum_{k' \in B(x)} \exp\left(q^S \cdot k' / \tau_1\right)} \tag{3}$$

In this formula, $P_1(X)$ denotes the positive set, and $B(X)$ represents the input batch of source domain samples. As suggested by work [34], a smaller τ_1 value leads to a more uniform feature distribution. Hence, we set τ_1 to 0.1 in the contrastive loss for the shared subspace.

During the training process, we start with self-supervised contrastive loss, followed by supervised contrastive loss. In the self-supervised scenario, $P(x)$ consists of the sample itself. In the supervised case, we focus on increasing the cosine similarity among points with the same label. Therefore, $P(x)$ is defined as the set of samples within the current batch $B(X)$ that have the same label as the ith sample, excluding the ith sample itself:

$$P(i) := \left\{ j \in B(X) \text{ s.t. } y_j^S = y_i^S \text{ and } j \neq i \right\} \tag{4}$$

3.4. Subject Space Procedure

The Subject Space Procedure follows the Common Space Procedure, involving the processing of all samples from both source and target domains. These samples undergo feature extraction via the Common Feature Extractor, resulting in shared physiological signal features, denoted as $\{q_i^S, y_i^S\}_{i=1}^{N_S}$ and $\{q^{T_i}\}_{i=1}^{N_T}$, from different subjects in response to identical emotional stimuli. The data from each source and target domain is then mapped into a unique feature subspace for each subject using a single-layer fully connected network, called the Subject-specific Feature Extractor (SFE). This mapping yields subject-specific features $\{Z_j^S, Y_j^S\}_{j=1}^{M_S}$ and $\{Z_j^T\}_{j=1}^{M_T}$.

To align the target domain features Z^T with the source domain features Z^S, the Maximum Mean Discrepancy (MMD) loss is used to estimate the feature distance between the target and source domains within the subject-specific feature subspace. The MMD loss helps align the feature distribution of the target domain with each source domain and is defined as:

$$L_{\mathrm{MMD}} = \left\| \frac{1}{M^S} \sum_{i=1}^{M^S} \Phi\left(\mathrm{X}_i^S\right) - \frac{1}{M^T} \sum_{j=1}^{M^T} \Phi\left(\mathrm{X}_j^T\right) \right\|_{\mathcal{H}}^2 \tag{5}$$

3.4.1. Inter-domain aggregation

Previous studies have indeed shown that MMD can achieve promising results in cross-subject emotion recognition [35]. However, the MMD is limited to minimizing the marginal distribution between source and target domains in the latent space, which we consider insufficient. To reinforce the consistency of categories between the target and source domains, we introduce contrastive learning within the subject-specific subspaces for inter-domain aggregation.

In experiments with video stimuli eliciting emotional responses, the label noise commonly arises from subjects' varied emotional reactions [17]. In typical supervised contrastive learning, the positive

pairs are composed of samples from the same class. However, treating samples from the same class as positive pairs may introduce noise, especially when those samples are associated with unreliable labels, thereby potentially compromising the integrity of the experiment. Such an approach could cause the model to overly generalize across different classes, thus blurring the distinctions between unique emotional samples. Therefore, to mitigate this issue, we opt to utilize source domain features to extract prototype embedding vectors μ_c for each class $c \in 1, 2, 3, \ldots, C$. These prototype embedding vectors serve as representative embeddings and are utilized to construct stable and reliable positive pairs. To update the prototypes, we adopt a moving average method:

$$\mu_c = \text{Normalize}(\gamma \mu_c + (1-\gamma) z), \quad \text{if } c, y = \arg\max(\text{SSC}(z)) \tag{6}$$

Here, μ_c is the momentum prototype for class c, calculated as the moving average of normalized query embeddings z that are correctly predicted to belong to class c. The hyperparameter γ is adjustable.

Utilizing the momentum prototype μ_c and the flexibility offered by the adjustable hyperparameter γ enables us to refine the stability and accuracy of our embeddings. This strategic implementation sets the stage for our next advancement: our contrastive learning approach, which is inspired by the innovative principles of the Mo-Co model [24]. The objective is to foster the consistency between the target and source domains concerning class labels by establishing a connection between information interactions in the target and source domains through contrastive learning. This occurs after achieving a similar data distribution in the subject-specific subspace via the MMD loss.

In our setup, we project q through the SFE to obtain z and bring it closer to class-conditioned prototype vectors μ_c within its subject-specific subspace. The early batches of z^S generated from the source domains are stored in a queue denoted by $\{z_j^S\}_{i=1}^J$. This strategy results in a substantial number of negative pairs (where the queue size J is significantly larger than the batch size N), thereby contributing to an enhanced representation of the context [36,37]. Due to the absence of label information for target domain samples, we use a classifier trained within each subject-specific subspace to predict the data of the target domain. Target domain samples with prediction scores exceeding a predefined threshold θ are assigned pseudo label $\hat{y}$. These samples, now labeled as $Z^K = \{z_i^T\}_{i=1}^K$, are paired with the extracted representations μ_c to form positive pairs $\{(z_i^T, \mu_c)\}_{i=1}^K$ within each subject-specific subspace. After each training step, all z^S from that batch are stored in a queue of size J. Therefore, for each domain, we have the following contrastive loss L_{con2}:

$$L_{con2} = -\frac{1}{K}\sum_{i=1}^{K} \log \frac{\exp(z_i^T \cdot \mu_c / \tau_2)}{\sum_{k' \in A(x)} \exp(z_i^T \cdot k' / \tau_2)} \tag{7}$$

Here, $A(x) = Z^K \cup \mu \cup queue$ includes all normalized embedding vectors. The hyperparameter τ_2 is set larger than τ_1 used in L_{con1} to facilitate contrastive learning in the subject-specific spaces.

3.5. Classification procedure

In the Classification Procedure of the MSCL model, we train subject-specific classifiers (SSC) using encoded features z from the SFE for result prediction. The SSC component consists of M individual softmax classifiers, each aligned with a particular subject-specific subspace. We implement the Generalized Cross Entropy (GCE) loss function [38] to address the challenge, due to the potential presence of noise labels in emotional data, particularly in experiments involving emotional responses to video stimuli. GCE is particularly effective in handling noisy labels as it prevents the model from overly focusing on samples with such labels during training. The classification loss using GCE is calculated as:

$$L_{cls} = \sum_{i=1}^{M} \frac{1 - (SSC_i(Z_i^S)^q)}{q} \tag{8}$$

Here, q is an adjustable hyperparameter, offering flexibility in the model's sensitivity to noisy labels.

Furthermore, due to the individual differences in brain activities elicited by the same emotional stimuli, the classifiers trained in different subject-specific subspaces may exhibit varied performance. A straightforward averaging of their prediction results may not be optimal. To address this issue, we assign weights to classifiers in each subject-specific subspace based on their similarity to the target domain. This weighted approach for predictions on target domain samples is given by:

$$\hat{Y}^T = \sum_{i=1}^{M_S} \omega_i \cdot SSC_i(Z_i^T) \tag{9}$$

The weight ω_i for each subject-specific subspace is determined by the proportion of positive pairs in that subspace relative to the mini-batch size:

$$\omega_i = \frac{K_i}{|B|} \tag{10}$$

3.6. Optimizing MSCL

The MSCL model is trained using both source and target domain EEG data, comprising M^S and M^T subjects, respectively, with a total of N^S and N^T samples. The training process involves several key steps, including Common Feature Extraction, Subject-specific Feature Mapping, Contrastive Learning for Class Information, Contrastive Learning for Class Information, Classification and GCE Loss.

Common Feature Extraction: The CFE projects all source and target domain data into a shared feature subspace for common feature extraction. Within this subspace, the contrastive loss L_{con1} is computed to ensure a balanced distribution of samples and to extract low-level common features.

Subject-specific Feature Mapping: The model employs N subject-specific subspaces to map the features of each subject. In these subspaces, the MMD loss $L_{\mathbf{MMD}}$ is computed for both source and target domains to capture domain-specific characteristics.

Contrastive Learning for Class Information: After the data distributions in the subject subspaces of source and target domains becoming sufficiently aligned (indicated by iteration $> \theta$), the contrastive loss L_{con2} is introduced to regulate the interaction between source and target domain samples with respect to class information.

Classification and GCE Loss: Subject-specific classifiers are employed for classification tasks, assessing effectiveness using the Generalized Cross Entropy loss $L_{\mathbf{GCE}}$.

The overall loss function L for training MSCL combines these different components:

$$L = L_{\mathbf{GCE}} + \alpha L_{\mathbf{MMD}} + \beta(L_{con1} + L_{con2}) \tag{11}$$

In this formulation, α and β are the adjustable parameters that balance the contribution of each loss component to the overall training process. The training of MSCL, as outlined in Algorithm 1, aims to optimize this combined loss function. By minimizing L_{con1}, the model ensures even distribution of samples in the common space, extracting essential low-level features. Minimizing $L_{\mathbf{MMD}}$ helps obtain domain-invariant features, aligning the data distribution of the source and target domains in the subject subspaces. Minimizing L_{con2} enhances the interaction of class information between the target and source domains. Lastly, minimizing $L_{\mathbf{GCE}}$ not only improves the classification accuracy of source domain data but also mitigates the impact of noise labels on the model.

Fig. 2. Video clip snapshots from the EEG-based emotion recognition experiment, depicting the stimuli for negative, neutral, and positive emotional responses.

Algorithm 1: Overview of MSCL

Input : Iteration T, source domain data $\{(X_i^S, Y_i^S)\}_i^{M_S}$ and target domain data $\{X^T\}$

1: **for** t = 1, ..., T do
2: Take n samples $\{x_j^{S_i}, y_j^{S_i}\}_{j=1}^n$ from source domains and $\{x_j^T\}_{j=1}^n$ from target domain
3: $\{q_j^{Si}\}_{j=1}^n, \{q^T\} \leftarrow CFE(\{x_j^{Si}, y_j^{Si}\}_{j=1}^n, \{x_j^T\}_{j=1}^n)$
4: $L_{con1} \leftarrow CFE$
5: $\{z_j^{Si}\}_{j=1}^n, \{z_j^T\}_{j=1}^n \leftarrow SFE(\{z_j^{Si}\}_j^n = 1, \{z^T\})$
6: $L_{\mathbf{MMD}}, L_{con2} \leftarrow SFE$
7: $\{\hat{y}_j^{Si}\}_{j=1}^n, \{\hat{y}_j^T\}_{j=1}^n \leftarrow SSC(\{z_j^{Si}\}_{j=1}^n, \{z_j^T\}_{j=1}^n)$
8: $L_{\mathbf{GCE}} \leftarrow SSC$
9: Update model by minimizing the total loss
10: **end for**
11: **return** $\{\hat{Y}^T\}$;

Output: Prediction of target domain data, $\{\hat{Y}^T\}$;

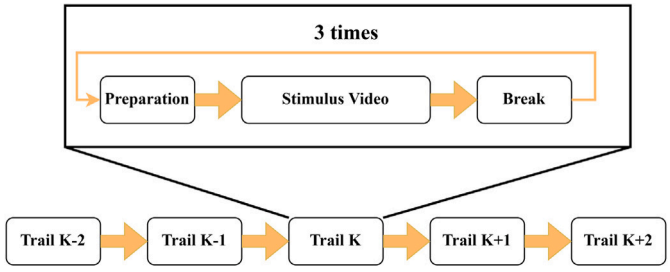


Fig. 3. Experimental flowchart illustrating the sequence of trials in the emotion recognition experiment. Each trial starts with a preparation phase where the subject gets ready for the experiment. This is followed by the presentation of a stimulus video, which is carefully selected to evoke specific emotional responses from the subject. After the video, there is a break period to allow the subject's emotional state to return to neutral, ensuring they are ready for the next trial without the previous one influencing their response. This cycle repeats across multiple trials (K − 2 to K + 2) to collect comprehensive EEG data for analysis.

4. Experiments and results

In this section, we will provide a comprehensive overview of the CEED Dataset, followed by a brief introduction to the SEED Dataset [7]. Subsequently, we will investigate the implementation details of our model. Finally, we will present the key findings, relevant studies, and visualization analysis.

4.1. Dataset

4.1.1. CEED Dataset

We recruit 50 college students, involving 40 males and 10 females with an average age of 21.9 years, all providing informed consent for participation. The experiment utilizes 15 video clips as stimuli, and is about three minutes long. There videos are categorized into three emotional states: negative, positive, and neutral, as shown in Fig. 2. Participants initiate the experiment by pressing a key upon reading the instructions. The experiment is structured into five segments, each beginning with the participant's initiation, followed by a 30-s rest period after each video. This structure includes a 5-s prompt, followed by a video clip designed to induce a specific emotional state, repeated three times for each of the three emotions, as illustrated in Fig. 3. After watching all the clips, participants complete a questionnaire about their emotional reactions. EEG signals are recorded using the BrainVision actiChamp-Plus system equipped with a 64-electrode acti-CAP, adhering to the international 10–20 system excluding the reference electrode, which results in data from 63 channels at a sampling rate of 1000 Hz, detailed in Fig. 4.

4.1.2. SEED Dataset

The SEED Dataset [7], a widely recognized benchmark for emotion recognition algorithms, includes EEG data from 15 subjects (8 females, average age 23.27 years, SD 2.37 years). Each subject watches 15 movie clips, which are carefully selected to elicit specific target emotions: positive, neutral, and negative, with five clips for each emotion category. The duration of each clip varies between 185 and 265 s, averaging 226 s. Subjects participate in three separate experimental sessions, spaced at least a week apart. In each session, they watch a movie clip, culminating in 15 trials. The EEG signals are captured using the ESI NeuroScan System 2, which features 62 channels positioned according

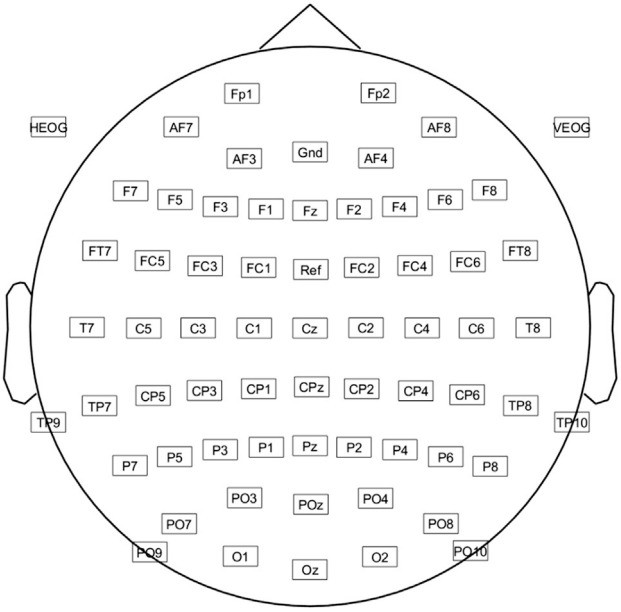


Fig. 4. Electrode placement schematic for the EEG cap based on the 64-channel acti-CAP snap standard used in the emotion recognition experiment. The diagram includes the HEOG (horizontal electrooculography) and VEOG (vertical electrooculography) electrodes for capturing eye movements, with the Oz electrode as the reference.

to the international 10–20 system and operates at a sampling rate of 1000 Hz.

4.2. Data preprocessing

For CEED Dataset, the initial EEG data undergo downsampling to 200 Hz and are subjected to bandpass filtering between 0.3 Hz to 50 Hz. Independent Component Analysis (ICA) is then applied to remove artifacts from eye movements, muscle activities, or environmental noise, adhering to a conservative criterion that eliminates only those independent components (ICs) exhibiting significant and persistent noise. Each trial's data is segmented into 1-second intervals. Differential Entropy (DE) features are extracted for predefined frequency bands (Delta: 1–3 Hz, Theta: 4–7 Hz, Alpha: 8–13 Hz, Beta: 14–30 Hz, and Gamma: 31–50 Hz), resulting in a total of 305 features per EEG segment (5 frequency bands × 61 channels). A Linear Dynamic System (LDS) method is utilized to smooth all features, accounting for temporal dependencies in emotional changes and filtering out noise-related EEG components unrelated to emotions [39].

For the publicly available SEED Dataset, which has already been downsampled to 200 Hz and filtered within the 0–75 Hz range, a similar denoising process is applied. A bandpass filter from 4 Hz to 50 Hz is applied, and the data is re-referenced to a common average reference. Regarding the CEED Dataset, DE features for the five frequency bands are extracted and smoothed using the LDS method, yielding 310 features per EEG segment (5 frequency bands × 62 channels).

4.3. Implementation details

The CEED Dataset comprises experimental data from 50 participants, resulting in 50 trials. When using the CEED Dataset as input, we randomly select data from 10 participants for analysis. For the SEED Dataset, we only choose data from the first period involving 15 participants to ensure ample the emotional information. Both datasets employ a leave-one-out evaluation method: for each participant, the MSCL model is assessed by treating them as the target domain and the other 14 participants as the source domain. The best-performing result is selected for each experiment. Although both datasets feature labeled samples, the labels of target domain samples are exclusively used for evaluation, not for training.

As outlined in Section 3, the MSCL model entails complexities across its three stages. Initially, for the CFE, a 3-layer MLP configuration (256, 128, and 64 nodes per layer) is chosen, with LeakyReLU layers acting as activation functions. In the Common Space Procedure, the data mapped through the CFE to the common space undergoes calculation of L_{con1} with τ_1 set at 0.1, using the unsupervised method early in training ($iteration < 400$). In the Subject-Specific Feature Extractor (SFE) and the Subject-Specific Classifier (SSC), dimensionality is reduced using linear functions: first from 64-D to 32-D, and subsequently to the number of classes, which is 3 for both datasets. After the linear layer in the SFE, a LeakyReLU activation function is applied. In contrast, the SSC employs a single linear layer without subsequent activation.

In the Subject Space Procedure, as data pairs map to subspaces via SFE, $L_{\mathbf{MMD}}$ is computed initially ($iteration < 420$), while L_{con2} remains uncalculated, essentially rendering L_{con2} zero. Class characterizations are computed and updated from the start of training, with an update rate γ set at 0.5. For L_{con2} calculation, target domain samples with predicted probabilities over 0.7 ($\theta > 0.7$) receive pseudo-labels for constructing L_{con2} loss, with τ_2 set at 0.3. The hyperparameter q for $L_{\mathbf{GCE}}$ computation is set at 0.55.

Training utilizes the Adam optimizer [40] with an initial learning rate of 0.01 over 100 epochs and a batch size of 128, meaning 128 samples from each domain per iteration. The model is trained using loss function L, with MMD as the distance metric between domains in feature space. All vectors involved in L_{con1} and L_{con2} computations are normalized. The coefficients α and β are dynamically adjusted, with α adapted to prioritize classification results and then achieve convergence between MMD and the classifier ($\alpha = \frac{2}{1+e^{-10*i/epoch}} - 1$). β is set to $\alpha/100$ in our experiments.

Table 2
Comparison results on SEED.

Dataset	Methods	Avg (%)	Std (%)
SEED	DAN	66.29	4.64
	MS-MDA	83.85	5.25
	DDA	78.14	10.71
	SeqCLR	85.77	–
	CLISA	86.40	6.4
	PCDG	87.3	0.021
	MSCL (Our)	88.18	7.55

Table 3
Comparison results on the CEED dataset.

Dataset	Methods	Avg (%)	Std (%)
CEED	DAN	46.15	4.59
	MS-MDA	89.08	4.56
	DDA	86.43	5.12
	MSCL (Our)	91.10	3.40

Table 4
Ablation study of MSCL on SEED of accuracy.

Dataset	Methods	Avg (%)	Std (%)
SEED	MSCL	88.18	7.55
	MSCL $w/o.L_{con1}$	80.22	6.92
	MSCL $w/o.L_{con2}$	86.50	9.64
	MSCL $w/o.L_{\mathbf{GCE}}$	86.76	8.15
	Baseline	79.90	10.77

4.4. Main results

To validate the effectiveness of our proposed MSCL model, we conduct a comparative analysis with several existing emotion recognition methods using both the CEED Dataset and the SEED Dataset. These methods encompass domain adaptation techniques such as DAN [41], MS-MDA [42] and DDA [43], as well as recently introduced contrastive learning methods like SeqCLR [30], CLISA [31] and PCDG [32], which have been applied in EEG-based emotion recognition. For the models DAN, MS-MDA, and DDA, we reproduce their results by utilizing their publicly available code. However, for models SeqCLR, CLISA and PCDG, which lack publicly available code, we cite the results reported in their papers. The summarized results are presented in Tables 2 and 3. The outcomes clearly indicate that our MSCL model consistently outperforms other baseline methods across various tasks, demonstrating significant improvements over the domain adaptation and contrastive learning models mentioned earlier, especially in the cross-subject scenario. Specifically, for the SEED Dataset, our MSCL model achieves a minimum improvement of 8% over the recently proposed domain adaptation models and a 2% improvement over the contrastive learning models. For the CEED Dataset, the MSCL model also exhibits superior performance, surpassing other comparison models by at least 2%. Additionally, the results highlight that our proposed method exhibits lower variance in predicting accuracy compared to previous domain adaptation methods, suggesting that our model possesses robustness. Figs. 5 and 6 depict the confusion matrices of the emotion recognition tasks in the cross-subject scenario on the SEED Dataset and the CEED Dataset, respectively. The model tends to misclassify negative emotions with the other two emotions, particularly neutral emotions. This observation is consistent with prior research narratives in the field [31,44,45].

4.5. Ablation studies

In order to better verify our model, we conduct an ablation study to understand the individual contributions of L_{con1}, L_{con2}, and the GCE to the MSCL model in detail. This study sequentially removes each of the three loss functions to assess their specific impacts on the model's

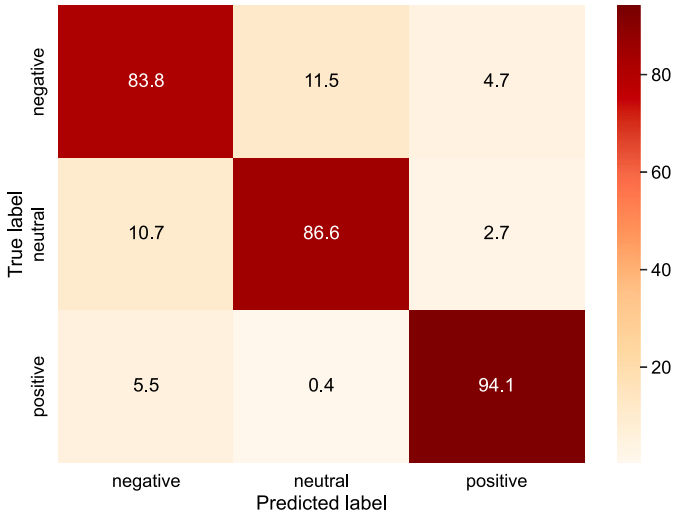


Fig. 5. The confusion matrix for the SEED Dataset.

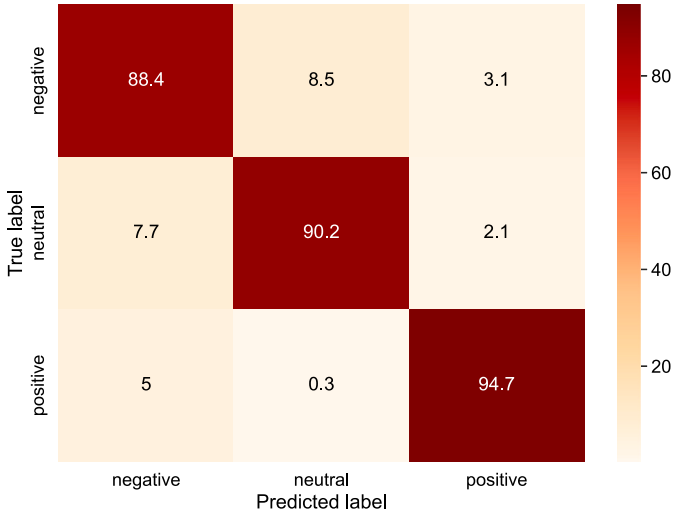


Fig. 6. The confusion matrix for The CEED Dataset.

performance. The findings from this ablation study are presented in Table 4.

When L_{con1} is excluded, identified as "$w/o.L_{con1}$" in the Table 4, the process is restricted to basic mapping within the common feature space. This reductionist approach, as shown in the second entry of the Table 4, precipitates a marked degradation in classification accuracy, thus highlighting the critical role of L_{con1} in extracting shared features and ensuring uniform data sample distribution among subjects.

The absence of L_{con2}, referred to as"w/o. L_{con2}", eliminates the comparative analysis of inter-class representations and the targeted alignment of domain representations, which, as evidenced in the third row of the table, negatively impacts the accuracy of emotion classification.

Choosing "$w/o.L_{GCE}$", which involves using conventional class cross-entropy (CE) instead of GCE, demonstrates the negative impact of noisy labels on training efficacy. However, the integration of GCE significantly mitigates these effects.

The baseline method, which discards L_{con1} and L_{con2} and replaces GCE with CE while retaining L_{MMD}, serves as a control to highlight the substantial improvements brought by the proposed loss functions. Collectively, these findings affirm that L_{con1}, L_{con2}, and L_{GCE} are crucial in enhancing the model's emotion recognition capabilities.

Table 5
Sensitivity analysis of α and β on MSCL model performance.

Dataset	β	Avg (%)	Std (%)
SEED	α	83.27	10.01
	$\alpha/10$	85.49	9.55
	$\alpha/100$	88.18	7.55
	$\alpha/1000$	83.20	8.15

Table 6
Sensitivity analysis of τ_1 and τ_2 on MSCL model accuracy.

τ_1	τ_2					
	0.3	0.4	0.5	0.6	0.7	0.8
0.05	85.46	82.26	81.34	81.25	80.82	79.73
0.1	**88.18**	85.89	84.34	83.43	80.93	79.07
0.15	86.91	86.13	83.34	85.45	81.23	80.82
0.2	83.57	84.53	83.46	84.87	81.62	79.12

4.6. Sensitivity analysis

To determine the influence of the hyperparameters on the MSCL model's efficacy, a series of experiments is designed in our research. These experiments scrutinize the ramifications of modifying the learning rate and the hyperparameter q in the context of the GCE. The learning rate and q are chosen from predefined ranges of 0.001 to 0.3 for the learning rate and 0.2 to 0.8 for q, ensuring that all other hyperparameters remain at their optimal levels. The resultant data, depicted in Fig. 7, encapsulate the cross-subject mean accuracy variations when applying the MSCL model to the SEED Dataset. While the model exhibits marginal sensitivity to variations in the learning rate and q, the overarching trend is one of stability. Observations indicate that increases in q are directly proportional to improvements in accuracy, alongside an expected the increase in training duration. These findings support the claims made in the referenced study [38].

The performance evaluation of the MSCL model is conducted by varying the hyperparameter γ from 0.3 to 0.8. The accuracy results, as shown in Fig. 8, demonstrate a relatively stable trend across different values of γ. Detailed observation reveals that the model achieves its highest accuracy when γ is set to 0.5. This optimal value suggests that an intermediate update rate for the momentum prototype provides the most effective representations for contrastive learning. At lower γ values (0.3 and 0.4), the model's accuracy is slightly reduced, indicating that the prototypes may be overly sensitive to recent data changes, leading to instability. Conversely, at higher γ values (0.7 and 0.8), a slight decrease in accuracy is observed, implying that the prototypes are not adapting quickly enough to new data, resulting in less effective learning. The figure indicate that while the MSCL model maintains robustness across the tested γ range, the peak performance is achieved at $\gamma = 0.5$. This balance ensures that the momentum prototypes are neither too volatile nor too static, thereby providing a reliable basis for enhancing the model's performance in emotion recognition tasks.

The sensitivity analysis of the hyper-parameters α and β on the MSCL model's performance, as shown in Table 5, reveals significant insights pertinent to the model's optimization. The experiments involved varying the β parameter while keeping α at different scales: α, $\alpha/10$, $\alpha/100$ and $\alpha/1000$. The results indicate that when β is set to α, the model achieves an average accuracy of 83.27% with a standard deviation of 10.01%, suggesting that the performance is relatively low with considerable variability. As β is reduced to $\alpha/10$, the average accuracy improves to 85.49% with a reduced standard deviation of 9.55%, indicating that the model becomes more stable and performs better. The optimal performance is observed when β is set to $\alpha/100$, where the model attains the highest average accuracy of 88.18% and the lowest standard deviation of 7.55%. This optimal setting suggests that β at $\alpha/100$ provides the best balance between effective learning dynamics and model stability. However, further reducing β to $\alpha/1000$

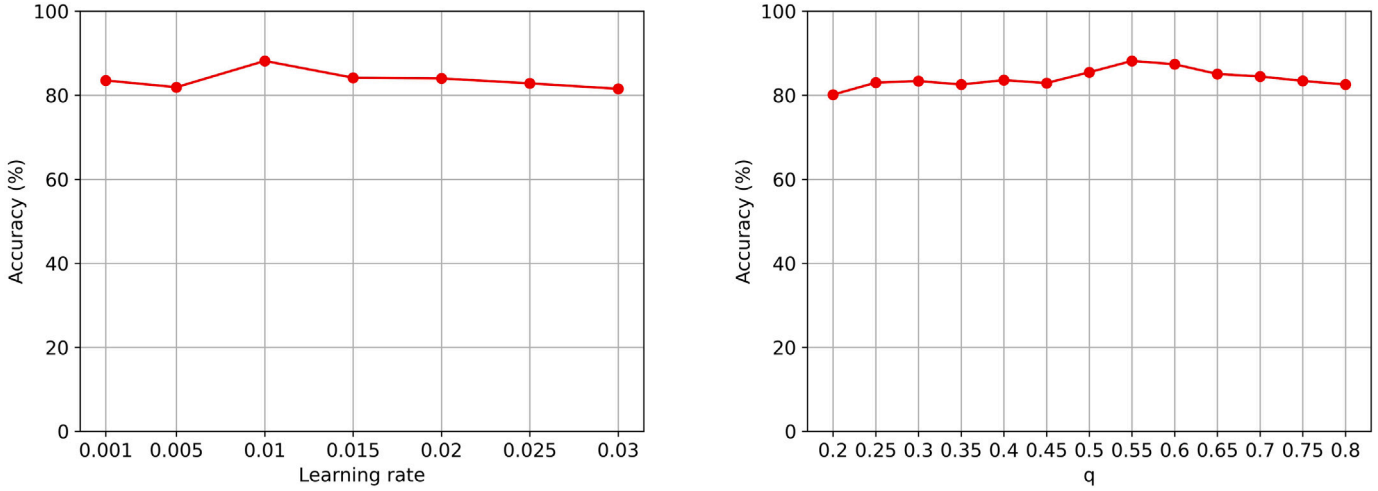


Fig. 7. The two graphs illustrate the effects of varying the learning rate (left) and the hyperparameter q (right) on the accuracy of the MSCL model.

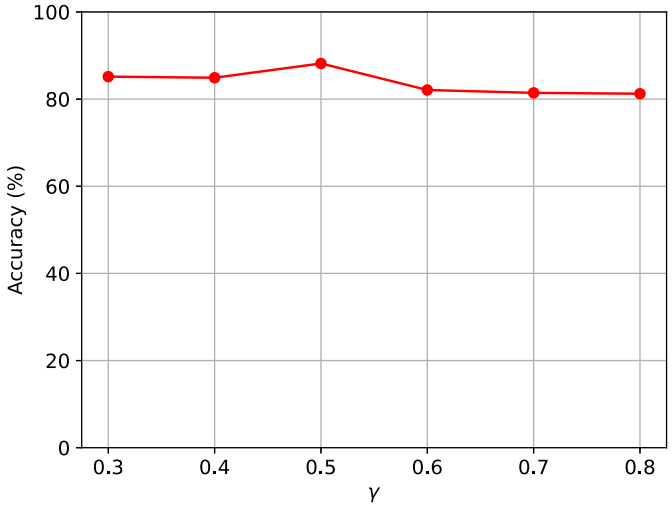


Fig. 8. The graphs illustrate the effects of varying the γ on the accuracy of the MSCL model.

results in a decrease in average accuracy to 83.20% and an increase in standard deviation to 8.15%, indicating that excessively small β values may impair the model's learning capacity, leading to reduced performance and stability.

The Table 6 presents the sensitivity analysis of the hyperparameters τ_1 and τ_2 on the accuracy of the MSCL model. The values of τ_1 and τ_2 range from 0.05 to 0.2 and 0.3 to 0.8, respectively. The results illustrate how these hyperparameters affect the model's performance in terms of accuracy. The analysis indicates that the highest accuracy of 88.18% is achieved when τ_1 is set to 0.1 and τ_2 is set to 0.3, suggesting that this combination provides the best performance for the model. Generally, a moderate value for τ_1 combined with a relatively lower τ_2 value is optimal. The model consistently performs well across different τ_2 values when τ_1 is set to 0.1, indicating robustness. However, increasing τ_1 beyond 0.1 leads to a decrease in accuracy, suggesting that higher values for τ_1 may not be optimal. Similarly, the performance is optimal when τ_2 is lower (around 0.3 to 0.4), with accuracy generally decreasing as τ_2 increases beyond 0.5, indicating that a lower τ_2 value is beneficial for maintaining a good distribution of similarity scores. The model's performance is relatively stable when τ_1 is 0.1 or 0.15, with minor fluctuations in accuracy across different τ_2 values. However, when τ_1 is set to 0.05 or 0.2, more pronounced drops in accuracy are observed, indicating less stability.

In an effort to delineate the influence of batch size on the MSCL model's performance, an assessment is conducted within a cross-subject framework using the SEED Dataset. The results of this assessment are graphically represented in Fig. 9. The investigation reveals subject-specific accuracy oscillations as a function of varying batch sizes for eight subjects. Despite these individual variances, the aggregate accuracy trend across all subjects demonstrates notable stability. Initially, an increase in batch size is correlated with a gradual enhancement in the MSCL model's accuracy. However, the improvement in performance plateaus and then decreases when the batch size exceeds a threshold of 128, indicating an optimal range for the batch size. Beyond this range, the model's effectiveness diminishes. This observation highlights the importance of carefully calibrating the batch size to ensure optimal performance of the model. The analysis of subject-specific accuracies indicates that while individual accuracies fluctuate, the overall trend remains consistent across different batch sizes. For batch sizes ranging from 16 to 128, the model's performance generally improves, peaking at 128. As the batch size continues to increase beyond this point, a decline in accuracy is observed. This suggests that smaller batch sizes may not provide sufficient data diversity for effective training, while excessively large batch sizes might lead to suboptimal gradient updates due to averaging effects.

To investigate the impact of the number of subjects utilized as source domains on the MSCL model's accuracy, we execute a series of emotion recognition experiments. These experiments systematically increase the source domain count from 1 to 14, targeting eight subjects. The results of these assessments are methodically illustrated in Fig. 10. The figure reveals that while individual subjects experience varying degrees of accuracy fluctuation with the expansion of the source domain pool, there is a general trend of gradual improvement in accuracy across all subjects. As the number of source domains increases, the model's accuracy generally improves, although there are minor fluctuations in individual subject performances. Initially, as the source domain count increases from 1 to 4, there is a significant improvement in accuracy for most subjects. This uptrend, however, experiences some variability between the 5 to 8 source domain range, suggesting that inter-subject variability might influence these patterns. Beyond 8 source domains, the accuracy realigns with the former trend of improvement, indicating that further augmentation of source domains continues to benefit the model's performance. The collective evidence indicates that enriching the source domain repository significantly enhances the precision of MSCL in emotion recognition.

4.7. Space visualization

To enhance the insight into the discriminative capability of our proposed MSCL model, we visualize the high-dimensional EEG signal

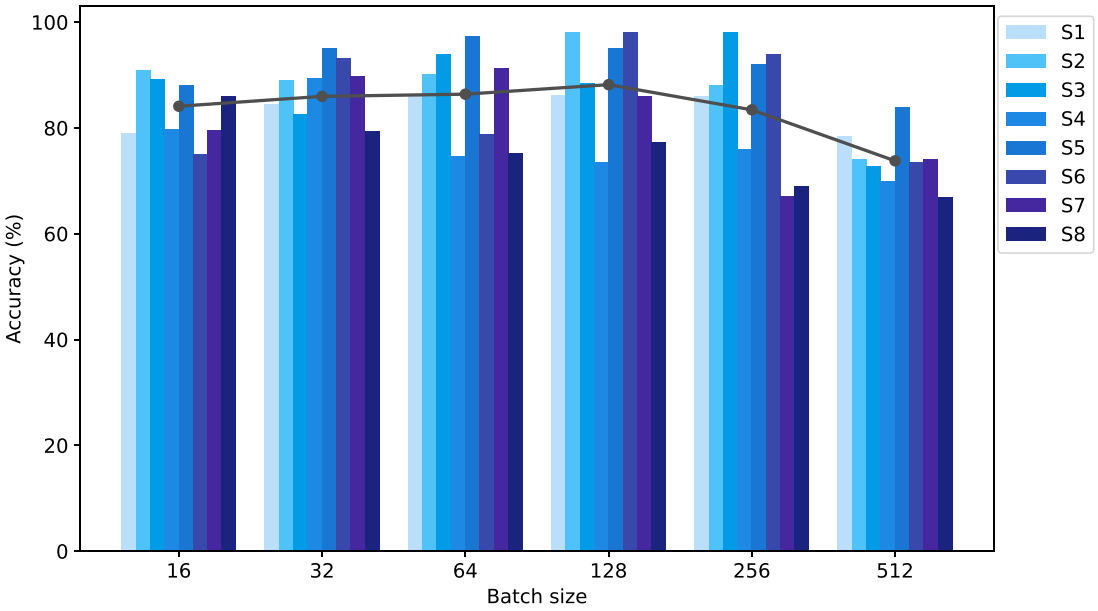


Fig. 9. The bar chart shows the accuracy variation of the MSCL model for four individual subjects (S1 to S8) with changing batch sizes. Each color corresponds to a different subject, with bars indicating individual accuracy. The line graph plots the mean accuracy across 14 subjects, revealing the overall trend in a cross-subject context.

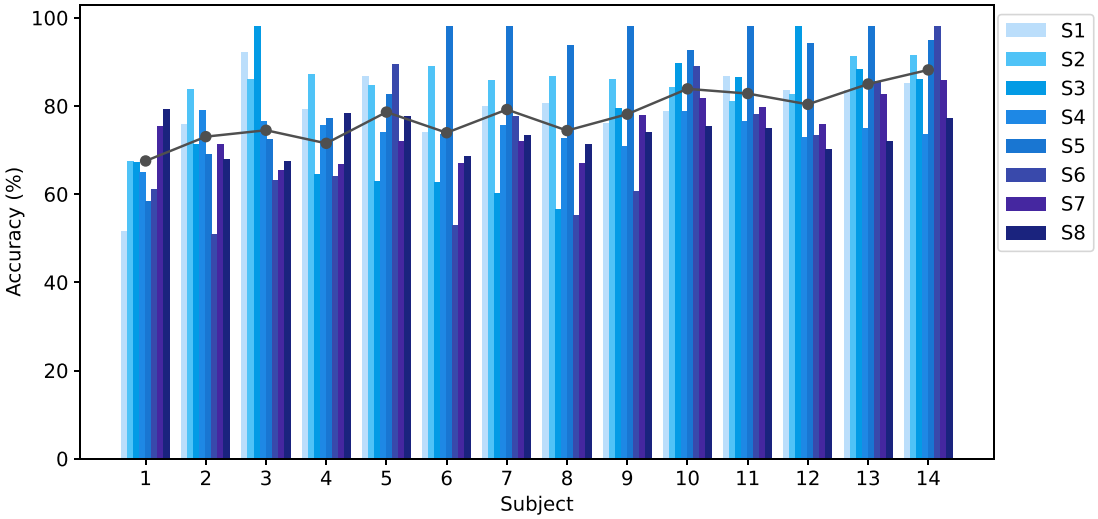


Fig. 10. The bar chart illustrates the accuracy of Subjects (S 1 to S 8) against the increasing number of source domains used in the MSCL model.

data using t-SNE [46], a tool adept at illustrating the complex data landscapes in a lower-dimensional space. We randomly select 100 EEG samples from each subject, corresponding to distinct source domains, and plot them in a cross-subject analysis on the SEED Dataset. In Fig. 11, the varied colors represent the different source domains, while the black signifies the target domain. To facilitate the clarity in the visualization, the target domain samples are rendered with transparency to mitigate visual clutter.

It is noteworthy that, as indicated in the top-left figure, the distribution of all EEG data from different subjects (represented by different colors) closely resembles each other. Most samples are concentrated within a particular region, with a few outliers observed in individual subjects. These distributions affirm the presence of shared characteristics in EEG data across all subjects, suggesting a slight overlap in their feature space.

As depicted in the top-right figure, different colored points in the common space exhibit a tendency to cluster, forming distinct clusters, which can be identified as three categories. Furthermore, the distribution of points in the common space appears to be relatively uniform. This observation strongly supports the efficacy of our L_{con1} in extracting shared features from EEG data across all subjects.

When the source domain data is combined with the target domain data into a single feature subspace, following the MSCL model's approach, the ideal outcome is depicted in the bottom plot. Here, the data points from the target domain should be closely aligned with those from the source domains. Additionally, the source domain should form discernible clusters, reflecting the model's sophisticated feature differentiation. The visualization confirms this, demonstrating our model's ability to smoothly blend and distinguish the neurophysiological markers of emotional states captured by EEG data.

5. Discussion

The proposed MSCL model in this work constitutes an advancement in the domain of emotion recognition from EEG data, particularly when addressing the complex cross-subject challenges. Different from traditional methodologies, MSCL adeptly integrates unsupervised and supervised contrastive learning techniques. This integration allows for

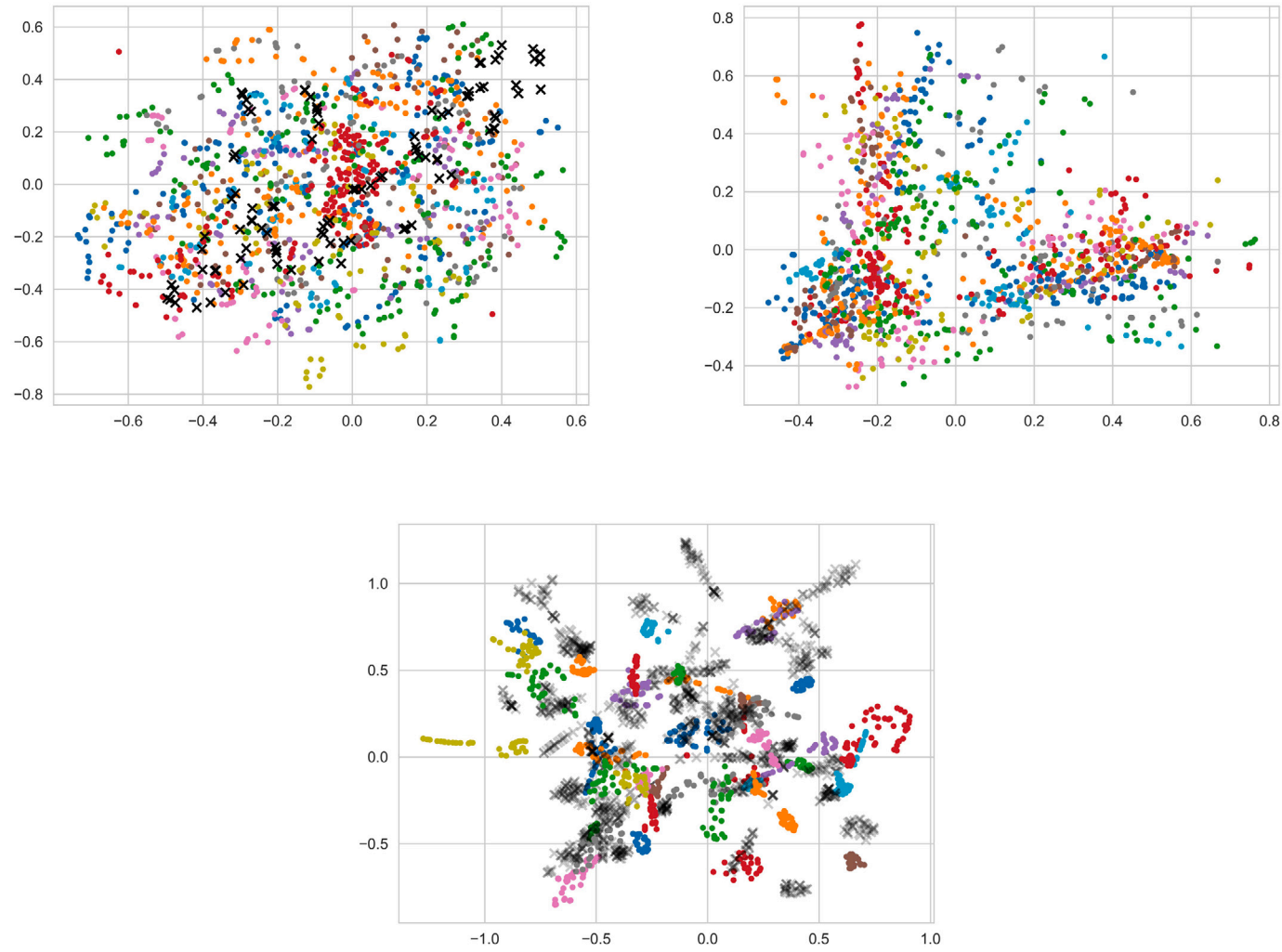


Fig. 11. Visualization of sample distributions from our study, captured via t-SNE. Each color and symbol represent samples from different subjects or experimental conditions. The top left plot indicates the initial unprocessed distribution, the top right shows samples after the common feature extraction phase, and the bottom plot reveals the refined clustering after the complete MSCL model processing. These plots illustrate the model's effectiveness in feature separation and space transformation for emotion classification tasks.

the extraction of both shared emotional features and subject-specific responses, thereby encapsulating a broader emotional spectrum. The novel dual contrastive learning approach, combined with the GCE, allows the MSCL to identify universal patterns across various subjects while maintaining the uniqueness of each individual's emotional experience. This capability marks a distinct improvement over previous models like MS-MDA, DDA, and CLISA, which primarily concentrated on aligning either the global distribution or class-specific features, but not both.

5.1. Achieving the best performance with the MSCL model

The comparative analysis, as shown in Tables 2 and 3, highlights the outstanding performance of the MSCL model in terms of accuracy and reduced variance across various subjects. This evidence supports MSCL's robustness and its broader applicability over the traditional domain adaptation and the contrastive learning approaches. Furthermore, the detailed ablation studies, as shown in Table 4, underscore the critical role of each component within the MSCL framework. Specifically, these studies highlight the indispensability of the contrastive learning elements for the model's refined interpretation of emotional data.

An examination of MSCL's responsiveness to changes in hyperparameters, particularly learning rates and the q parameter in the GCE, reveals the importance of precise adjustments in these areas. Our research indicates that although MSCL maintains stable performance across a wide range of hyperparameter settings, to identify an optimal configuration can enhance accuracy while minimizing computational resources. Additionally, visual analyses provided in Fig. 11 clarify potential areas of ambiguity within the model, notably between negative and neutral emotional states. These insights lay the foundation for focused improvements in subsequent versions of the model, with the objective of enhancing its ability to distinguish between different emotional states and increasing its overall effectiveness.

5.2. Complexity and time efficiency analysis

5.2.1. Model complexity analysis

The MSCL exhibits various levels of computational complexity across its components. The CFE, which comprises a MLP with 256, 128, and 64 nodes in each layer respectively, forms the basis of the feature extraction process. Assuming the input feature dimension is (d), the computational complexity for processing through these layers can be expressed as $O(d \cdot 256 + 256 \cdot 128 + 128 \cdot 64)$. Therefore, the total complexity simplifies to $O(d \cdot 256)$, highlighting that the CFE's complexity is primarily dependent on the input feature dimension.

For the inter-class contrastive learning component, the L_{con1} calculation involves determining the similarity between samples within a batch. Given a batch size of B, the complexity of computing the

contrastive loss for each sample is $O(B)$. Since there are B samples in each batch, the overall complexity becomes $O(B^2)$, indicating that this step scales quadratically with the batch size.

The inter-domain contrastive learning and domain adaptation steps involve more intricate computations. The SFE, which processes both source and target domain samples, has a complexity of $O(f \cdot 32)$ per sample, where f is the feature dimension. Given N_s source samples and N_t target samples, the total complexity is $O((N_s+N_t)\cdot f \cdot 32)$. Additionally, the L_{MMD} calculation, which aligns the feature distributions between source and target domains, has a complexity of $O(N \cdot M)$, where N and M are the numbers of source and target samples respectively. The L_{con2} in the subject-specific feature space, which involves comparing each target sample with K class prototypes, adds a complexity of $O(N_t \cdot K)$.

In the classification procedure, the complexity of classifying each sample is $O(32 \cdot 3)$, simplifying to $O(96)$. For a batch of B samples, this results in a total complexity of $O(B \cdot 96)$, indicating a linear relationship with the batch size.

5.2.2. Time efficiency analysis

Analyzing the time efficiency of the MSCL model involves considering the computational complexity of its various components.

During the feature extraction phase, the time complexity of the CFE is primarily influenced by the input feature dimension d and the batch size B. The overall time complexity, $O(B \cdot d \cdot 256)$, suggests that the computation time increases linearly with both the batch size and the input dimension, maintaining manageable computational demands even as these parameters grow.

For contrastive learning, the inter-class contrastive learning step's quadratic complexity, $O(B^2)$, implies that the computation time will increase significantly with larger batch sizes. This step could become a bottleneck if the batch size is not appropriately managed. Similarly, the inter-domain contrastive learning steps involve multiple complex calculations. The overall complexity of $O((N_s+N_t)\cdot f \cdot 32+N_t \cdot K+N \cdot M)$ highlights that the time efficiency in these steps is sensitive to the number of source and target samples, as well as the feature dimension.

In the classification procedure, the process benefits from a linear complexity with respect to the batch size, $O(B \cdot 96)$, indicating efficient scaling with larger batches.

5.3. Limitations

While the MSCL model represents a important progress, the increase in training time with the addition of more source domains highlights the necessity for careful sample selection. It is crucial to explore the training methodologies that improve efficiency while maintaining the model's high level of performance. Additionally, although the GCE method effectively addresses the issue of noisy labels, the investigation of more advanced techniques for noise mitigation is essential to further strengthen the model's robustness. Furthermore, the task of establishing a potent negative sample repository for the contrastive learning poses a call for creative strategies, such as the employment of data augmentation tactics, to broaden the model's experiential spectrum.

6. Conclusion

In this research, we introduce a multi-source representation selection transfer learning model based on the contrastive learning. This model enhances the extraction of common features in EEG data and addresses inter-subject differences through contrastive learning. Additionally, we also incorporate the GCE to mitigate the negative impact of noise labels in training and improve the model's robustness. We evaluate the performance of the proposed model in the cross-subject scenario based on the SEED Dataset and the CEED Dataset, and compare it with the existing state-of-the-art methods. Extensive experimental results demonstrate that the MSCL achieves the best results on both datasets, highlighting its advantages in addressing individual differences and noise label issues in BCI systems. In the future work, we will continue our research by exploring the selection of more efficient model encoders, optimizing the construction of negative sample pools, and enhancing our datasets among other aspects.

CRediT authorship contribution statement

Xin Deng: Writing – review & editing, Supervision, Resources, Project administration, Investigation, Conceptualization. **Chenhui Li:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Xinyi Hong:** Writing – review & editing, Data curation. **Huaxiang Huo:** Writing – review & editing, Data curation. **Hongxing Qin:** Resources, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hongxing Qin reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 2272071, 61772097].

References

[1] E. Marg, Descartes'error: emotion, reason, and the human brain, Optom. Vis. Sci. 72 (11) (1995) 847–848.
[2] W. Li, W. Huan, B. Hou, Y. Tian, Z. Zhang, A. Song, Can emotion be transferred?—A review on transfer learning for EEG-based emotion recognition, IEEE Trans. Cogn. Dev. Syst. 14 (3) (2021) 833–846.
[3] W. Hu, G. Huang, L. Li, L. Zhang, Z. Zhang, Z. Liang, Video-triggered EEG-emotion public databases and current methods: a survey, Brain Sci. Adv. 6 (3) (2020) 255–287.
[4] Y. Ding, X. Hu, Z. Xia, Y.-J. Liu, D. Zhang, Inter-brain EEG feature extraction and analysis for continuous implicit emotion tagging during video watching, IEEE Trans. Affect. Comput. 12 (1) (2018) 92–102.
[5] X. Hu, J. Chen, F. Wang, D. Zhang, Ten challenges for EEG-based affective computing, Brain Sci. Adv. 5 (1) (2019) 1–20.
[6] Y. Xin, A. Qv, Matlab simulation analysis of power spectrum estimation based on welch method, Mod. Electron. Tech. 33 (2010) 7–9.
[7] R.-N. Duan, J.-Y. Zhu, B.-L. Lu, Differential entropy feature for EEG-based emotion classification, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering, NER, IEEE, 2013, pp. 81–84.
[8] X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, Z. Zhao, N. Kumar, P. Marttinen, EEG based emotion recognition: A tutorial and review, ACM Comput. Surv. 55 (4) (2022) 1–57.
[9] M.-y. Zhong, Q.-y. Yang, Y. Liu, B.-y. Zhen, B.-b. Xie, et al., EEG emotion recognition based on TQWT-features and hybrid convolutional recurrent neural network, Biomed. Signal Process. Control 79 (2023) 104211.
[10] L. Yang, Q. Zhang, S. Chao, D. Liu, X. Yuan, Greedy-mrmr: An emotion recognition algorithm based on eeg using greedy algorithm, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2022, pp. 1329–1336.
[11] Y.-X. Yang, Z.-K. Gao, X.-M. Wang, Y.-L. Li, J.-W. Han, N. Marwan, J. Kurths, A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG, Chaos 28 (8) (2018).
[12] S. Tripathi, S. Acharya, R. Sharma, S. Mittal, S. Bhattacharya, Using deep and convolutional neural networks for accurate emotion classification on DEAP data, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, 2017, pp. 4746–4752.

[13] M. Ye, C.P. Chen, T. Zhang, Hierarchical dynamic graph convolutional network with interpretability for EEG-based emotion recognition, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[14] Y. Wei, Y. Liu, C. Li, J. Cheng, R. Song, X. Chen, TC-net: A transformer capsule network for EEG-based emotion recognition, Comput. Biol. Med. 152 (2023) 106463.
[15] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao, Y.-D. Zhang, EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network, Knowl.-Based Syst. 265 (2023) 110372.
[16] S. Katsigiannis, N. Ramzan, DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices, IEEE J. Biomed. Health Inform. 22 (1) (2017) 98–107.
[17] C. Li, Y. Hou, R. Song, J. Cheng, Y. Liu, X. Chen, Multi-channel EEG-based emotion recognition in the presence of noisy labels, Sci. China Inf. Sci. 65 (4) (2022) 140405.
[18] R. Zhou, Z. Zhang, H. Fu, L. Zhang, L. Li, G. Huang, Y. Dong, F. Li, X. Yang, Z. Liang, PR-PL: A novel transfer learning framework with prototypical representation based pairwise learning for EEG-based emotion recognition, 2022, arXiv preprint arXiv:2202.06509.
[19] Z. He, Y. Zhong, J. Pan, An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition, Comput. Biol. Med. 141 (2022) 105048.
[20] J. Pan, R. Liang, Z. He, J. Li, Y. Liang, X. Zhou, Y. He, Y. Li, ST-SCGNN: A spatio-temporal self-constructing graph neural network for cross-subject EEG-based emotion recognition and consciousness detection, IEEE J. Biomed. Health Inf. (2023).
[21] T. Song, S. Liu, W. Zheng, Y. Zong, Z. Cui, Y. Li, X. Zhou, Variational instance-adaptive graph for EEG emotion recognition, IEEE Trans. Affect. Comput. 14 (1) (2021) 343–356.
[22] Y. Li, J. Chen, F. Li, B. Fu, H. Wu, Y. Ji, Y. Zhou, Y. Niu, G. Shi, W. Zheng, GMSS: Graph-based multi-task self-supervised learning for EEG emotion recognition, IEEE Trans. Affect. Comput. (2022).
[23] J. Li, H. Hua, Z. Xu, L. Shu, X. Xu, F. Kuang, S. Wu, Cross-subject EEG emotion recognition combined with connectivity features and meta-transfer learning, Comput. Biol. Med. 145 (2022) 105519.
[24] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9729–9738.
[25] Z. Wu, Y. Xiong, S.X. Yu, D. Lin, Unsupervised feature learning via non-parametric instance discrimination, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3733–3742.
[26] M. Ye, X. Zhang, P.C. Yuen, S.-F. Chang, Unsupervised embedding learning via invariant and spreading instance feature, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6210–6219.
[27] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: Beyond empirical risk minimization, 2017, arXiv preprint arXiv:1710.09412.
[28] K. Wickstrøm, M. Kampffmeyer, K.Ø. Mikalsen, R. Jenssen, Mixing up contrastive learning: Self-supervised representation learning for time series, Pattern Recognit. Lett. 155 (2022) 54–61.
[29] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.
[30] M.N. Mohsenvand, M.R. Izadi, P. Maes, Contrastive representation learning for electroencephalogram classification, in: Machine Learning for Health, PMLR, 2020, pp. 238–253.
[31] X. Shen, X. Liu, X. Hu, D. Zhang, S. Song, Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition, IEEE Trans. Affect. Comput. (2022).
[32] H. Cai, J. Pan, Two-phase prototypical contrastive domain generalization for cross-subject EEG-based emotion recognition, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2023, pp. 1–5.
[33] M. Soleymani, S. Asghari-Esfeden, Y. Fu, M. Pantic, Analysis of EEG signals and facial expressions for continuous emotion detection, IEEE Trans. Affect. Comput. 7 (1) (2015) 17–28.
[34] F. Wang, H. Liu, Understanding the behaviour of contrastive loss, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2495–2504.
[35] G. Bao, N. Zhuang, L. Tong, B. Yan, J. Shu, L. Wang, Y. Zeng, Z. Shen, Two-level domain adaptation neural network for EEG-based emotion recognition, Front. Hum. Neurosci. 14 (2021) 605246.
[36] P. Bachman, R.D. Hjelm, W. Buchwalter, Learning representations by maximizing mutual information across views, Adv. Neural Inf. Process. Syst. 32 (2019).
[37] Y. Tian, D. Krishnan, P. Isola, Contrastive multiview coding, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16, Springer, 2020, pp. 776–794.
[38] Z. Zhang, M. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels, Adv. Neural Inf. Process. Syst. 31 (2018).
[39] L.-C. Shi, B.-L. Lu, Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 6587–6590.
[40] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[41] H. Li, Y.-M. Jin, W.-L. Zheng, B.-L. Lu, Cross-subject emotion recognition using deep adaptation networks, in: Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25, Springer, 2018, pp. 403–413.
[42] H. Chen, M. Jin, Z. Li, C. Fan, J. Li, H. He, MS-MDA: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition, Front. Neurosci. 15 (2021) 778488.
[43] Z. Li, E. Zhu, M. Jin, C. Fan, H. He, T. Cai, J. Li, Dynamic domain adaptation for class-aware cross-subject and cross-session EEG emotion recognition, IEEE J. Biomed. Health Inf. 26 (12) (2022) 5964–5973.
[44] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, T. Song, A novel bi-hemispheric discrepancy model for EEG emotion recognition, IEEE Trans. Cogn. Dev. Syst. 13 (2) (2020) 354–367.
[45] P. Zhong, D. Wang, C. Miao, EEG-based emotion recognition using regularized graph neural networks, IEEE Trans. Affect. Comput. 13 (3) (2020) 1290–1301.
[46] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).